RICE UNIVERSITY

Metabolic engineering approaches to biosynthesize terpenoids in *Saccharomyces cerevisiae*

by

Caroline V. McNeil

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Seiichi P. T. Matsuda, Professor,
Department Chair, Department of Chemistry
Department of Biochemistry and Cell Biology

Ronald J. Parry, Professor
Department of Chemistry
Department of Biochemistry and Cell Biology

George Bennett, Professor
Department of Biochemistry and Cell Biology

HOUSTON, TEXAS

July 2008

July 2008

ABSTRACT

Metabolic engineering approaches to biosynthesize terpenoids in *Saccharomyces cerevisiae*

by

Caroline V. McNeil

Terpenoids are the largest class of natural products and are typically isolated from natural sources. However, heterologous expression of terpene synthases in microbial hosts such as *E. coli* or *Saccharomyces cerevisiae* has become an attractive alternative. *S. cerevisiae* has an intact sterol biosynthetic pathway, and many of the intermediates also serve as precursors for terpene synthases. Metabolic engineering efforts focus on optimizing product yields through increasing carbon flux through the desired pathway, removing competing enzymes, or altering enzymatic activity.

This work describes the metabolic engineering of *S. cerevisiae* to enhance terpene production by exploiting these three approaches. Diterpene synthases were expressed in a yeast strain previously reported to accumulate the diterpene precursor geranylgeranyl pyrophosphate (GGPP). The strains produced milligram amounts of GGPP hydrolysis products geranylgeraniol and geranyllinalool, as well as the GGPP cyclization products *ent*-copalyl pyrophosphate, *ent*-kaurene, and abietadiene. Because diterpene production is limited by transit peptides targeting diterpene synthases into plastids, protein expression was increased by co-expressing a chloroplast processing enzyme in two different

diterpene-producing strains. The in vivo-generated mature diterpene synthases functioned more effectively, thereby increasing cyclization yield.

This thesis also describes a new method for controlling farnesyl pyrophosphate (FPP) hydrolysis product profile by adjusting media pH. Hydrolysis was found to be partially controlled by a phosphatase DPP1, however a majority of FPP hydrolysis is non-enzymatic. In a squalene synthase deletion strain, FPP accumulates and hydrolyzes readily to farnesol and nerolidol, and the ratios of these products are determined by the pH of the media.

Finally, a yeast strain was constructed to increase production of the 30-carbon triterpene precursors oxidosqualene (OS) and dioxidosqualene (DOS) by over-expressing the sterol biosynthesis rate-limiting enzyme 3-hydroxy-3-methylglutaryl CoA reductase (HMG1) in a lanosterol synthase deletion background. This strain accumulated twenty times more OS and DOS than the strain with only the native HMG1. Over-expression of squalene epoxidase (ERG1) in a lanosterol synthase background greatly enhanced the levels of DOS compared to OS.

Acknowledgments

This thesis would not have come about without support from my amazing families. First, the biggest thank you to my lab, who has been like family these last few years. My thesis advisor, Prof. Seiichi P. T. Matsuda welcomed me into his lab and taught me an incredible amount of molecular biology and chemistry. He has given me so many amazing opportunities, both in terms of research projects and teaching experiences, and I will forever be grateful to him helping me become who I am today. Dr. William K. Wilson has helped me with NMR experiments and GC-MS repairs, and he has always been available to give professional and personal advice. Dr. Hui Shan provided invaluable GC-MS help and provided a very friendly work environment. Dr. Silvia Lodeiro has provided a lot of valuable advice as she became a good friend. Dr. Uttam Dasgupta, our friendly dentist, has shown me his incredible strength in times of adversity. Prof. Quanbo Xiong taught me a lot about analysis, molecular biology, and patience, and he has been an incredible friend from the first day I came to this lab.

Former lab members Dr. Renee LeClair and Dr. Gia Fazio helped me with molecular biology and always came through when I needed help in or out of lab. Dr. Mariya Kolesnikova shared nearly every moment of graduate school with me. Her friendship is like no other, and I am truly grateful for all the advice, supportive shoulders, and fun experiences she has shared me through these years. Current graduate students Pietro Morlacchi and Dorianne Castillo-Rivera have brought fresh perspectives into our lab and I have really enjoyed working with them.

Alyssa Baevich is an incredible undergraduate I feel blessed to have worked with, as she brings laughter, hard work, and an incredible sweetness to everything she does. A

large part of this thesis is because of her help. Kaylah Dunbar has only worked with me a short time but I am very proud of the scientist she is becoming and all she has and will accomplish. Allie Obermeyer and Carl Onak shared space with me for many years and I am proud to know them as they go on to their next adventures. David Lynch, McKenzie Smith, Diana Sttiven, and Aparna Bhaduri have all helped me realize how much I love teaching and have worked hard to get where they are today.

To my thesis committee members, thank you for helping me with this thesis. Prof. Ronald Parry has listened to nearly every one of my chemistry seminars and has provided valuable feedback and advice through the years. I also thank Prof. George Bennett for kindly serving on my committee and providing valuable feedback.

My family has been with me every step of the way, and I would not be here today without them. First my amazing husband, Chase McNeil, has been my rock and my everything, showing endless patience and understanding through graduate school. His love and support have gotten me through some tough times, and I could not have come this far without him.

My father, Tom Pardue, always encouraged me to pursue my dreams and provided cleaning supplies and snacks whenever I have needed them. My sister, Tammy, has always been there as a listening ear and a supportive shoulder, and her husband Jean-Christophe has been interested in what I have to say and encourages me to bigger things. My brother Jim shares his stories of adventure that keep me in tune with the outside world. My grandmothers, Maxine Pardue and Charlotte West, have always believed in me and helped in ways too big to measure. Their cookies, hugs, and calls kept me afloat when I needed it most.

My second family has also been a great source of love and support. Mark and Terri McNeil have been wonderful in-laws, and their support through graduate school has been appreciated in ways I can never put into words. My sister-in-law Tanis has been a great source of encouragement and always keeps me in reality when I lose track of it.

This thesis is in loving memory of my grandfather James West, who passed away in August 2004. The loving support and encouragement he and my grandmother provided my last years of college allowed me to succeed in being the person they believed I could be. I would also like to acknowledge my grandfather Prentiss Pardue who left this earth in December 2003. He was an amazing person and taught me a lot about life and hard work. This is also in memory of my dear friend Kristin Holt, who became a wonderful friend in college and encouraged me to study chemistry even when times got tough.

This thesis is a special memoriam to my mother, Virginia Pardue. Without her love, encouragement, and patience, I would not be the person I am today. Her death was a source of sadness but also showed me what truly matters in this world. She taught me faith, and hope, and those two intangible objects are what have brought me to where I am today.

To all who have encouraged me, believed in me, and helped me this thesis is a testament to a lifetime of family, friends, and amazing teachers. And if not for my faith which gave me strength when I was weak, I would never have come this far.

To my loving family

In loving memory of
C. Virginia Pardue

Table of Contents

List of Figures

List of Tables

Chapter 1

Background and Introduction

The largest class of natural products is the terpenoids, a diverse group of compounds with many different functions in plants, fungi, and animals. Terpenoids can be divided into two main groups – primary metabolites and secondary metabolites. Within the primary metabolites are the sterols, such as cholesterol, which make up cellular membranes, and hormones, such as estrogen, testosterone, or the plant hormones gibberellins. The majority of terpenoid structural diversity arises from secondary metabolites isolated primarily from plants. These compounds serve a variety of functions in plants, including defending against insects and counteracting stress situations or infections. In addition to being utilized by plants, a large number of terpenes also have commercial applications. These include medicinal applications such as the chemotherapeutic agent paclitaxel (trade name Taxol®) and the anti-malarial drug artemisinin, or fragrance applications such as patchouli or nootkatone, two components in many perfumes.

Terpenes are derived from the five-carbon isoprene unit and differ in their number of carbons, size and number of rings, and functionality. Though typically these compounds are derived from the mevalonate pathway, beginning with acetyl-CoA to the 5-carbon precursor isopentenyl pyrophosphate (IPP), a mevalonate-independent pathway was discovered in 1993 in bacteria, green alga, and higher plants.[1-3] The non-mevalonate pathway begins with glyceraldehyde-3-phosphate and pyruvate and ultimately forms IPP

and later farnesyl pyrophosphate (FPP). This thesis focuses on the mevalonate pathway, so the non-mevalonate pathway will not be discussed in depth.

The mevalonate pathway generates the necessary substrates for terpene production, including sterol biosynthesis. Plants, fungi, and mammals all produce sterols, and their biosynthetic pathways are identical up until production of the 30-carbon oxidosqualene (OS). The pathways diverge after this point to form the membrane sterols required by each organism. Plants cyclize OS to cycloartenol which is further metabolized to sitosterol and stigmasterol. Fungi and mammals cyclize OS to lanosterol, and the pathways then diverge to form cholesterol in mammals and ergosterol in fungi. This thesis focuses on manipulation of the sterol biosynthetic pathway of *Saccharomyces cerevisiae* for biosynthesis of many different terpenoids.

S. cerevisiae sterol biosynthetic pathway

Sterol biosynthesis starts with the conversion of acetyl-CoA to mevalonate in a two-step enzymatic process (Figure 1.1). First, two molecules of acetyl-CoA condense to form acetoacetyl-CoA via a Claisen condensation mechanism, catalyzed by acetoacetyl-CoA thiolase (ERG10).[4,5] 3-Hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) synthase (HMGS, ERG13)[4,6] catalyzes an aldol condensation of another molecule of acetyl-CoA with acetoacetyl-CoA to yield HMG-CoA.

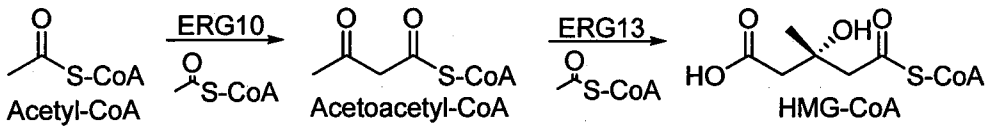

Figure 1.1. The first two steps of the sterol biosynthetic pathway involve condensation of three molecules of acetyl-CoA to form HMG-CoA. The reactions are catalyzed by ERG10 and ERG13.

In yeast with low levels of ergosterol produced, activities of both ERG10 and ERG13 increased substantially, while enzymatic activity is much lower when excess ergosterol is present.[7] As ERG10 activity increased, ERG13 activity increased at the same rate, suggesting ERG13 activity is dependent on that of ERG10 and that the two enzymes are regulated by similar mechanisms.

HMG-CoA reductase (HMGR) is the rate-limiting enzyme in sterol biosynthesis,[8] catalyzing the formation of mevalonic acid from HMG-CoA (Figure 1.2).[9] *S. cerevisiae* encodes two genes *HMG1* and *HMG2* that both possess HMGR activity.[10] The N-termini (339 amino acids) of these two enzymes are anchored to the endoplasmic reticulum (ER), while the C-termini reach into the cytoplasm and contain the enzyme active sites.[11]

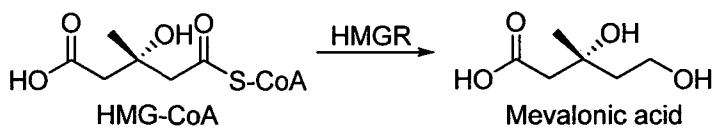

Figure 1.2. Reduction of HMG-CoA to mevalonic acid by HMG-CoA reductase.

These two enzymes are highly regulated by several factors, including feedback control[12,13] and HMGR degradation.[14] HMG1 was found to be fairly stable while HMG2 was quickly degraded.[14] When ERG10, ERG13, or even HMGR activity was repressed,

production of the HMGR protein was increased, suggesting that an intermediate downstream of HMGR was involved in feedback regulation.[13]

HMG1 and HMG2 are ER-bound proteins and therefore subject to ER degradation. Degradation of these two enzymes is controlled by flux within the mevalonate pathway; increased degradation was observed when flux increased and less degradation was recorded as flux decreased.[14] HMG2 is ubiquitinated by the UBC7 enzyme, with ubiquitination dependent on flux through the mevalonate pathway, and this mechanism is what regulates HMG2 stability and degradation in yeast.[15] FPP is the feedback signal that controls ubiquitination and therefore degradation of HMG2.[15,16]

Overexpression of HMG1 in yeast led to formation of karmellae, stacked membranes situated about the nucleus, of the endoplasmic reticulum whereas overexpression of HMG2 did not produce these structures.[17] A 552-residue truncation (trHMG1) was constructed that contained only the catalytically active C-terminal region by removing the N-terminal membrane-spanning domain.[18] This modification formed a soluble protein and removed transcriptional regulation. Overexpression of this trHMG1 dramatically increased squalene accumulation but only modest improvement in sterol production was observed.[18,19] This result indicated that another rate-limiting enzyme must exist after squalene biosynthesis.

Mevalonic acid, or its deprotonated form mevalonate, is converted to the 5-carbon terpene precursor IPP in three ATP-dependent steps (Figure 1.3). First, mevalonate is phosphorylated, catalyzed by mevalonate-5-phosphotransferase (ERG12), to yield phosphomevalonate.[20] Mevalonate-5-pyrophosphate forms from a second phosphorylation catalyzed by phosphomevalonate kinase (ERG8).[20,21] Finally,

mevalonate-5-pyrophosphate is decarboxylated by pyrophosphomevalonate decarboxylase (ERG19), yielding IPP.[22,23]

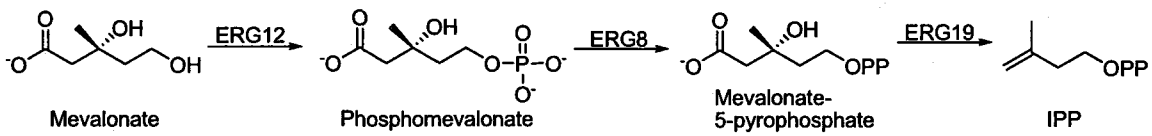

Figure 1.3. Conversion of mevalonate to IPP. Two phosphorylations and a decarboxylation reaction yield the 5-carbon isoprene unit.

IPP is utilized in chain elongation to form the sterol and terpene precursors geranyl pyrophosphate (GPP), FPP, and geranylgeranyl pyrophosphate (GGPP). To form GPP, a molecule of IPP condenses with a molecule of its isomer dimethylallyl pyrophosphate (DMAPP). Isomerization of IPP to DMAPP is carried out by IPP isomerase (IDI1),[24] forming the more energetically favorable internal trisubstituted double bond (Figure 1.4).

Figure 1.4. Isomerization of IPP to DMAPP, catalyzed by IPP isomerase.

The next three steps in sterol biosynthesis involve a family of enzymes known as prenyltransferases. These enzymes catalyze condensations with IPP, DMAPP, and other polyprenyl intermediates in either head-to-tail or head-to-head condensation reactions. These enzymes are involved in chain-length elongation and play a vital role in sterol, and terpenoid, biosynthesis. Prenyltransferases contain a conserved DDXXD motif[25] to bind

the allylic substrates,[26] and prenyltransferases such as FPP synthase and GGPP synthase also possess two aspartate-rich regions.[25]

FPP synthase[27] (ERG20) catalyzes two separate condensation reactions in *S. cerevisiae*. First, ERG20 catalyzes the head-to-tail condensation of one molecule of DMAPP and one molecule of IPP to form the 10-carbon GPP. GPP does not accumulate but instead condenses with another molecule of IPP to form FPP.

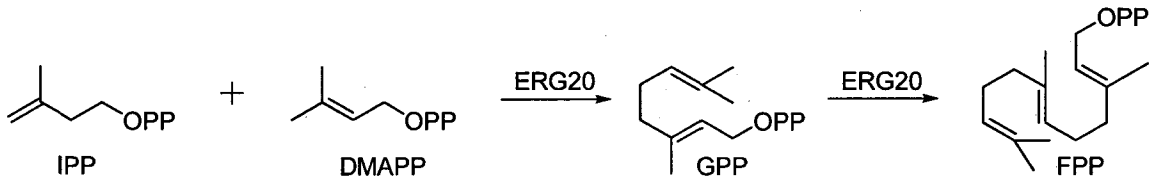

Figure 1.5. Condensation to GPP and FPP by FPP synthase (ERG20).

Wild-type *S. cerevisiae* does not produce geraniol or farnesol,[28] hydrolysis products of GPP and FPP, respectively. Through a series of mutagenesis experiments two yeast mutants were discovered that lacked squalene synthase activity. One of these mutants secreted farnesol into the medium, while the other mutant secreted both farnesol and geraniol.[28] It was determined that an ERG20 mutant Lys197Glu led to formation of geraniol and linalool into the media, suggesting that this residue was vital for FPP synthase activity.[29] Blanchard and Karst proposed that the lysine residue was important for substrate binding, and replacing this residue with a glutamic acid residue decreased the affinity for GPP and therefore reduced chain-length elongation. The catalysis to yield FPP was not completely eradicated, leading to formation of farnesol in addition to the monoterpene alcohols.

Mutagenesis experiments undertaken on avian FPP synthase revealed that two other mutations, Ala114Trp and Asn144Trp, led to smaller active site pockets that no longer accepted GPP as substrate.[30] These mutants only accepted DMAPP and IPP, therefore functioning only as GPP synthases.

FPP is a branch point in the pathway and is utilized in several biosynthetic pathways. In addition to sterol biosynthesis, FPP is involved in protein prenylation and the formation of GGPP, dolichols, and ubiquinones.

Because FPP is the substrate for multiple pathways in yeast, ERG20 is a highly-regulated enzyme. To determine the effect of ERG20 on sterol biosynthesis, however, the enzyme cannot be completely deleted. Therefore mutagenesis experiments to lower FPP synthase activity have been undertaken.[31] Lowering FPP synthase activity 42% (compared to wild-type) did not cause a significant drop in ergosterol biosynthesis, but decreasing activity to 20% reduced ergosterol production by 50%. When enzymatic activity was lowered further, ergosterol biosynthesis did not change appreciably, most likely because FPP was being diverted from other pathways to sustain sterol metabolism for growth. Because of this diversion, the yeast did not grow as quickly or as strongly as observed in earlier strains.[31]

Alternatively, overexpression of FPP synthase led to a 6-fold increase in FPP synthase activity and a 20% increase in HMGR activity.[32] The overexpression also increased dolichol and ergosterol production. Under aerobic conditions excess FPP is diverted to dolichol production, and these results suggest that ERG20 serves a role in regulating FPP to either squalene or dolichol production. Any ergosterol surplus can be stored as acyl esters.[33]

Yeast utilizes GGPP primarily as the substrate for protein prenylation. GGPP synthase (BTS1) catalyzes the condensation of FPP with another molecule of IPP to form the 20-carbon molecule.[34] Because demand for GGPP in yeast is modest, little GGPP accumulates in wild-type *S. cerevisiae*.[35]

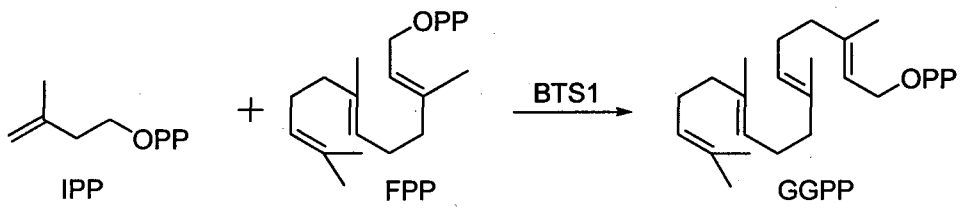

Figure 1.6. FPP and IPP condense to form GGPP.

Protein prenylation is catalyzed by three main families of enzymes, protein farnesyltransferase (PFTase) and two classes of protein geranylgeranyltransferase (PGGTaseI and PGGTaseII).[36] PFTase and PGGTaseI catalyze the alkylation of a cysteine residue within a CAAX motif (A is an aliphatic amino acid) at the C-terminus of various proteins. For farnesylation to occur, X must be Ala, Met, Gln, or Ser), while the geranylgeranyl moiety is added when X is Leu, Asn, or sometimes Phe.[37] The farnesylated or geranylgeranylated proteins have increased affinity for membranes, and the alkylation aids in membrane recognition.[36]

The final condensation reaction involves squalene synthase (ERG9) catalyzing the condensation of two FPP molecules to form the 30-carbon squalene (Figure 1.7).[38,39] This is the first committed step in sterol biosynthesis and requires magnesium ions and either NADH or NADPH for catalysis to occur. Unlike the FPP and GGPP synthases, squalene

synthase catalyzes a head-to-head condensation by first dimerizing FPP to presqualene pyrophosphate and then promoting a reductive elimination to form squalene.[40]

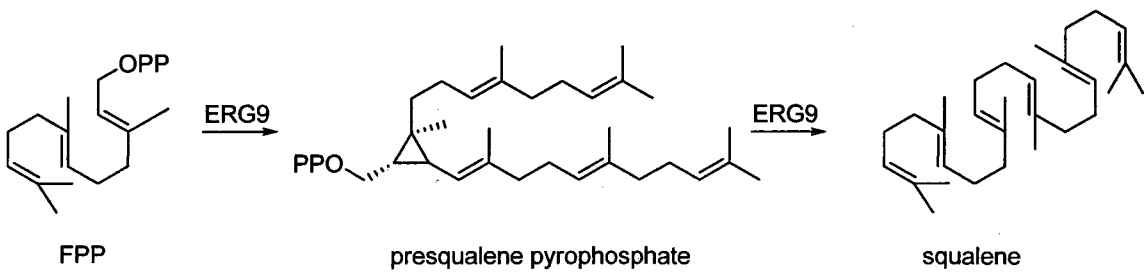

Figure 1.7. ERG9 catalyzes the condensation of 2 FPP molecules to form squalene. Presqualene pyrophosphate forms as an intermediate.

Yeast strains deficient in ERG9 activity accumulate FPP[41,42] and the FPP derivative farnesol, a hydrolysis product secreted into the medium.[28] Dolichol production also increased. ERG9 thus regulates FPP flux through sterol or dolichol biosynthesis.

Squalene epoxidase (ERG1) oxidizes squalene to (3*S*)-oxidosqualene[43,44] in a FAD- and NADPH-dependent reaction (Figure 1.8).[45] This is the first step of the pathway requiring molecular oxygen,[45,46] and strains grown in anaerobic conditions to block squalene epoxidase activity were found to accumulate squalene and no sterols.[47] ERG1 is under transcriptional regulation, with enzymatic activity dependent on ergosterol and lanosterol levels,[48] and the highest enzymatic activity was reported in strains deficient in lanosterol synthase activity.[49] When lanosterol biosynthesis is limited, OS can re-enter the ERG1 active site and form (3*S*, 23*S*)-dioxidosqualene (DOS) from epoxidation at the distal terminus.[50]

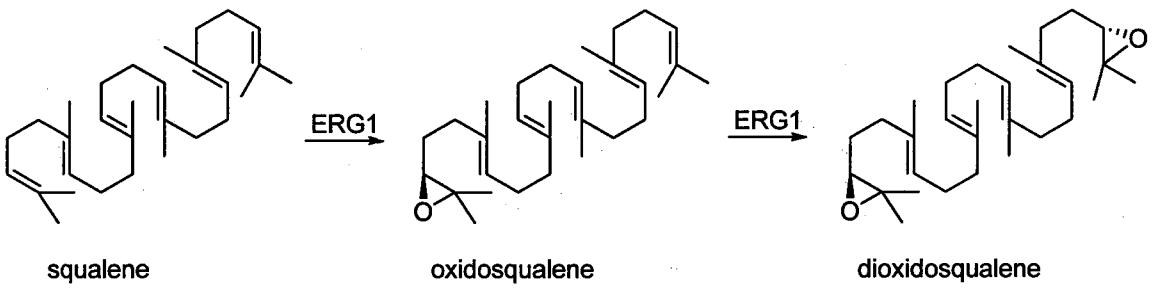

Figure 1.8. Epoxidation of squalene to form OS. OS can re-enter the ERG1 active site, yielding DOS.

Wild-type yeast grown in aerobic conditions accumulates almost no OS, which is efficiently cyclized to lanosterol. Overexpression of ERG1 in a trHMG1 overexpression background increased lanosterol production almost 4-fold and overall sterol yields nearly 2-fold.[51] Squalene accumulation simultaneously decreased 50%. OS did not accumulate in this strain; lanosterol and its metabolites were still the only products observed.

Lanosterol synthase (ERG7) cyclizes OS to the tetracyclic lanosterol through a cationic mechanism (Figure 1.9).[52] Lanosterol synthase can also accept dioxidosqualene to form epoxylanosterol,[53,54] a precursor for oxysterols that help regulate sterol biosynthesis.[55]

A series of 11 enzymes modify the lanosterol skeleton to the membrane ergosterol. First, lanosterol C-14 demethylase (ERG11) removes the C-14 methyl group from lanosterol in a P_{450}-dependent oxidation reaction.[56] The new double formed by ERG11 is reduced by Δ14-reductase (ERG24) to give 4,4-dimethyl zymosterol in an NADPH-dependent reaction.[57,58] A series of enzymatic reactions catalyzed by ERG25,[59] ERG26,[60] ERG27,[61] and ERG28[62] remove the two methyl groups at C-4, yielding zymosterol. C24-Methyltransferase (ERG6),[63] an isomerase (ERG2),[64,65] two desaturases

(ERG3 and ERG5),[66,67] and a reductase (ERG4)[68] act in succession to form the final ergosterol product.

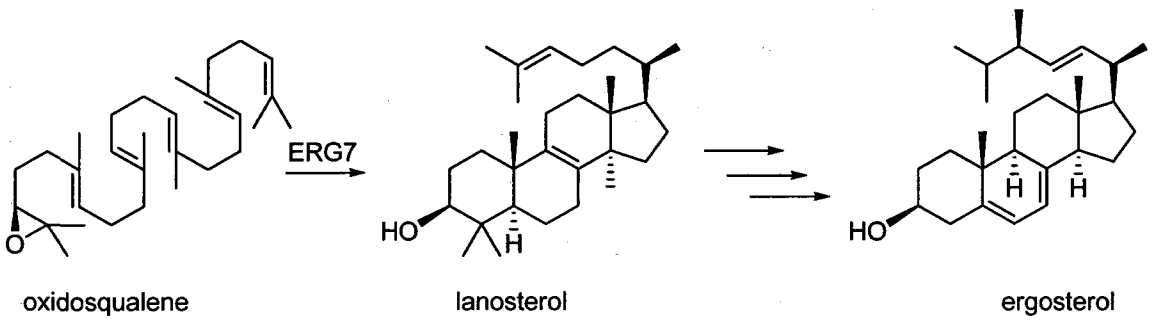

Figure 1.9. ERG7 catalyzes the cyclization of OS to lanosterol. Further modifications of lanosterol are carried out by 11 other enzymes to form ergosterol.

Other *S. cerevisiae* enzymes aiding in terpenoid biosynthesis

The UPC2 transcription factor is part of a superfamily of DNA binding proteins[69] that help regulate processes in *S. cerevisiae*. A single point mutation, *upc2-1*,[70] increased sterol uptake 5-fold compared to wild-type yeast under aerobic conditions.[71] The Gly888Glu mutant[70] also led to an increase in carbon flux through the sterol biosynthetic pathway, however the exact mechanism for this increase is still not understood.[72]

Without an intact sterol biosynthetic pathway, yeast needs to import sterols in order to survive. Because many steps in sterol biosynthesis require O_2, anaerobically grown yeast must import sterol. However, *S. cerevisiae* grown under aerobic conditions does not normally import exogenous sterols, even when mutations or inhibitors prohibit sterol biosynthesis. Simple sterol biosynthesis mutants are consequently not viable under aerobic growth conditions. To overcome this obstacle, sterol biosynthetic mutants included a heme biosynthetic mutant, which mimics anaerobic conditions, allowing yeast

to utilize exogenous sterol. A HEM1 deletion has been reported in an ERG7 deletion strain that allowed import of sterols under aerobic conditions.[73]

As noted above, yeast without a functional squalene synthase accumulate the FPP hydrolysis product farnesol. This hydrolysis has been attributed to phosphatases in the yeast. Two phosphatases, diacylglycerol pyrophosphate phosphatase (DPP1)[74] and lipid phosphate phosphatase (LPP1),[75] have been implicated in isoprenoid pyrophosphate hydrolysis.[76] Deletion of LPP1 led to 25% less hydrolysis activity against several substrates, including FPP and GGPP. DPP1 deletion showed a more dramatic reduction in hydrolysis activity, with 75% less activity recorded. The double deletion (*lpp1Δdpp1Δ*) showed almost no hydrolytic activity remained.[76] DPP1 showed strong hydrolysis activity against GGPP, with GGPP levels decreasing at the same levels geranylgeraniol (a GGPP hydrolysis product) increased. The double deletion strain showed a small amount of Mg^{2+}-dependent activity remained, which suggested that another phosphatase could be present in yeast to hydrolyze isoprenoid pyrophosphates. Indeed, a soluble form of an alkaline phosphate from yeast converted FPP to farnesol in vitro.[77]

Metabolic engineering for terpenoid production

Extraction of terpenoids from their natural sources is the most direct route, however a large amount of material is necessary to obtain enough material for analysis or commercial applications. Total synthesis eliminates the need for natural sources but has limitations due to lengthy, complex reaction sequences and low yields. Additionally, hazardous reagents are often utilized, making many of these syntheses unhealthy for the

environment. An attractive alternative is heterologous expression of terpene synthases in plants or microbial hosts.

Metabolic engineering of plants is a promising alternative to synthesis and extraction. As plants produce the majority of terpenoids found in nature, they already contain the necessary enzymes for precursor production. Work on transgenic plants to improve terpenoid production has focused on two main areas, increasing precursor availability and expressing foreign synthases or overexpressing native enzymes. However, metabolic engineering in plants is quite difficult because of poorer transformation efficiency, long generation times, insufficient understanding of the regulation of the biosynthetic pathways, and complexities that arise from the different subcellular locations of the plant enzymes.

Other challenges arise from rerouting precursors away from the plant's primary metabolism. For example, phytoene synthase was overexpressed in tomatoes, and even though carotenoid production increased, plant growth was stunted.[78] This stunting was attributed to diversion of the phytoene precursor GGPP from the biosynthesis of gibberellins, which are essential for plant growth. Similar results were reported when Besumbes et al. expressed taxadiene synthase, the diterpene synthase involved in paclitaxel biosynthesis, in *Arabidopsis thaliana*. Taxadiene yields were approximately 20 ng/g plant, but plant growth was also stunted, and this phenotype was also attributed to a foreign enzyme affecting regulation of enzymes in isoprenoid biosynthesis.[79] This amount of taxadiene shows the potential of this technology for terpene biosynthesis; however this method is still limited.

Microbial hosts utilized for terpenoid biosynthesis have demonstrated much higher titers than those reported in plants. *Escherichia coli* and *Saccharomyces cerevisiae* are two hosts widely reported to produce many classes of terpenes, including monoterpenes,[80,81] sesquiterpenes,[82-85] diterpenes,[86,87] triterpenes, and carotenoids. There are several advantages to using *E. coli* or *S. cerevisiae*, including fast doubling times (1-2 hours), well-known genetics, and ease of manipulation. *E. coli* has the disadvantage, however, of not having an endogenous sterol biosynthetic pathway, so many genes for terpene biosynthesis must be introduced into the strains.

However, *E. coli* has the non-mevalonate pathway to produce other terpenoids necessary for growth. Manipulation of this pathway in combination with expression of several terpene synthases has led to increased levels of monoterpenes,[81] sesquiterpenes, and diterpenes[81] in vivo. However these yields have not been as substantial as those reported in yeast.

S. cerevisiae, on the other hand, has an intact sterol pathway, producing the substrates necessary for terpene biosynthesis. To increase the amounts of intermediates, two main approaches are utilized. First, carbon flux is increased by overexpressing rate-limiting enzymes such as trHMG1. Removing or repressing competing enzymes has also enhanced terpene yields, with combinations of the two manipulations demonstrating even higher yields.

The 10-carbon monoterpenoids are derived from hydrolysis or cyclization of the precursor GPP and comprise a large number of flavors and fragrances. Further metabolism, such as oxidation, leads to compounds such as menthol and camphor. These compounds are produced mainly in plants, and engineering efforts in microbial hosts are

less common. *S. cerevisiae* does not naturally accumulate GPP and therefore does not produce monoterpenes. An *S. cerevisiae* mutant strain was previously reported to produce geraniol and linalool when an FPP synthase Lys197Glu mutation was expressed.[29] This mutant forms GPP preferentially over FPP. Geraniol synthase from *Ocimum basilicum* was expressed in wild-type and mutant *S. cerevisiae* (containing the FPP synthase mutant), and nearly 1 mg/L of geraniol and linalool were isolated from the mutant strain.[80]

Metabolic engineering to increase sesquiterpene production in microbial hosts has made great strides, with levels increased dramatically compared to what was first reported. Sesquiterpenes are produced from cyclization or hydrolysis of FPP by sesquiterpene synthases to a wide variety of hydrocarbon skeletons. In wild-type *S. cerevisiae* FPP does not accumulate, as it serves as a precursor in multiple biosynthetic pathways. The sesquiterpene synthase epicedrol synthase was expressed in wild-type and engineered yeast, and little epicedrol was isolated from the medium.[82] Overexpression of a soluble form of HMG1 (trHMG1) in a *upc2-1* background increased epicedrol yields 3-fold in comparison to that observed in wild-type strains. Overexpression of FPP synthase did not significantly increase sesquiterpene yields.[82]

Repression of the competing enzyme squalene synthase (ERG9) dramatically increased yields of several cyclization products. A down-regulated ERG9, in combination with an overexpressed trHMG1 and the *upc2-1* mutation, was utilized to increase artemisinic acid production in vivo. Amorphadiene synthase (the sesquiterpene synthase involved in artemisinin production), a P_{450} oxidase, and a P_{450} reductase were co-

expressed in this yeast strain, and up to 115 mg/L of artemisinic acid were produced in optimized culture conditions.[88]

Takahashi and coworkers also reported co-expression of a sesquiterpene synthase and a P_{450} oxidase in an engineered yeast strain to produce capsidiol yields over 50 mg/L.[84] This engineered strain included the overexpressed trHMG1 and deleted ERG9, but the phosphatase DPP1 was also deleted to see if this would decrease FPP hydrolysis to farnesol. Though farnesol levels decreased in the *dpp1Δ* strain, the manipulation did not significantly increase cyclized product yields.

These results demonstrate the utility of *S. cerevisiae* for sesquiterpene production, as well as the potential for accumulating oxidized terpenoids in vivo. Introducing sesquiterpene synthases in combination with P_{450} oxidases has shown high efficiency turnover to produce compounds such as artemisinic acid or capsidiol.

Diterpenoids, derived from the C-20 precursor GGPP, encompass a large class of natural products with diverse applications. Among these compounds are casbene, an antimicrobial and antifungal agent; paclitaxel, a drug used for breast cancer treatment; aphidocolin, an anti-viral agent; and *ent*-kaurene, the biological precursor for gibberellin plant hormones. Metabolic engineering of strains for diterpene production has been explored less than sesquiterpene production; however a few trends have been discovered that increased diterpene yields significantly.

The 20-carbon precursor GGPP is cyclized by a diterpene synthase to the final diterpene product or hydrolyzed to form geranylgeraniol or geranyllinalool. *E. coli* has been engineered to heterologously express several diterpene synthases, including *Abies grandis* abietadiene synthase,[87] *Ricinus communis* casbene synthase,[81] *ent*-kaurene

synthase,[81] and *Taxus brevifolia* taxadiene synthase,[86] which catalyzes the first step in paclitaxel biosynthesis. In one instance, the *E. coli* 1-deoxy-D-xylulose-5-phosphate synthase, a *Haematococcus pluvialis* IPP isomerase, and an *E. coli* FPP synthase carrying a mutation to produce GGPP were overexpressed in *E. coli*. Casbene synthase was expressed in this strain, and approximately 30 μg/L casbene was isolated.[81] Despite the overexpression of several enzymes, the diterpene titers were rather modest.

Wild-type *S. cerevisiae* produces little GGPP, as farnesyl pyrophosphate is diverted to other pathways. By overexpressing the yeast GGPP synthase BTS1 and a truncated HMG-CoA reductase, the amount of GGPP produced increased significantly.[35] Eight genes from *Taxus brevifolia* involved in paclitaxel biosynthesis, including the *Taxus* GGPP synthase, were introduced into wild-type *S. cerevisiae* and produced milligram amounts of oxidized diterpenoids.[89] This work demonstrates the potential of yeast for diterpene production, as much higher levels were observed than reported in *E. coli.*

One limitation of diterpene biosynthesis in microbial hosts is that diterpene synthases contain transit peptides that target the proteins to plastids within the plant.[90] The transit peptides are cleaved once inside the organelle to yield the mature protein,[91] but *E. coli* and *S. cerevisiae* do not have the mechanisms to cleave these transit peptides. To overcome this, diterpene synthases have been manually truncated to remove putative transit peptides, and this has increased diterpene yields in both *E. coli* and yeast.[87,92-96]

To study triterpene biosynthesis, yeast hosts were constructed to accumulate OS, the precursor for triterpene synthases. Yeast preferentially cyclizes OS to lanosterol for use in sterol biosynthesis, so in order to accumulate OS, the competing enzyme ERG7

was deleted. This strain, SMY8,[97] accumulates both OS and DOS and has allowed for characterization of numerous oxidosqualene cyclases.[52,98-102] This strain has also aided in characterization of DOS cyclization products by oxidosqualene cyclases, leading to discovery of novel enzymatic activity and novel compounds.[103] To improve yields, further work is necessary to increase accumulation of the OS precursor. As this thesis was being written, several strains engineered for triterpene biosynthesis were published.[104] Kirby and coworkers first overexpressed trHMG1 in wild-type yeast with *Artemisia annua* β-amyrin synthase, leading to an increase in squalene accumulation but not in β-amyrin yields. A combination of repressed ERG7 and trHMG1 overexpression doubled β-amyrin production from 3 mg/L (wild-type) to 6 mg/L.

Metabolic engineering of microbial hosts has become an important tool in the characterization and overproduction of a wide variety of terpenoids. Through a combination of increased metabolic flux and deletion of competing enzymes, the levels of terpenes produced in these engineered strains has increased dramatically, in some cases up to 30 mg/L of terpenes. Co-expression of terpene synthases and P_{450}-dependent oxidases has uncovered a wealth of new natural products, and this technology can be utilized to build libraries of potential medicines, perfumes, or other compounds with potential commercial value. Despite all the metabolic engineering work done thus far a great deal remains to be done to increase yields further.

Summary

The work presented here shows the utility of *S. cerevisiae* for terpenoid production. Manipulation of the sterol biosynthetic pathway allowed for increased

accumulation of several terpene substrates, including FPP, GGPP, OS, and DOS. First, a strain previously reported to accumulate the diterpene precursor GGPP was utilized to produce the gibberellin precursors *ent*-copalyl pyrophosphate and *ent*-kaurene. To further increase diterpene yields, an *Arabidopsis* chloroplast processing enzyme was co-expressed with the *Abies grandis* abietadiene synthase and with the *Arabidopsis ent*-copalyl pyrophosphate synthase and *ent*-kaurene synthase. These manipulations led to dramatic increases in diterpene production, as the signaling sequences were removed to solubilize the proteins, optimizing enzymatic activity. Additionally, *S. cerevisiae* was engineered to increase FPP accumulation for sesquiterpene production, resulting in high yields of the FPP hydrolysis products farnesol and nerolidol. In attempts to reduce hydrolysis, two phosphatases were deleted from the yeast and their effects were studied. Finally, triterpene substrate product profiles were influenced by overexpression of squalene epoxidase or the sterol rate-limiting enzyme trHMG1. These manipulations increased OS and DOS accumulation in yeast.

Chapter 2

Production of gibberellin precursors in metabolically engineered yeast

Introduction

Diterpenoids are a large class of natural products that serve diverse functions within plants, such as defense chemicals against insect herbivory or infection or as primary metabolites such as gibberellins, hormones that regulate growth, development, and other processes within the plant.[105] Many of the biological activities are adaptable for pharmaceutical purposes, and these diverse compounds have been extracted from plant sources for use as medicines such as paclitaxel, a chemotherapeutic from the Pacific yew tree,[106] and casbene, an antimicrobial and antifungal agent isolated from castor beans.[107,108]

However, extraction from natural sources has severe limitations, as large amounts of plants are needed to obtain sufficient amounts of material. Microbial hosts have become an attractive alternative for diterpene production and have been utilized to heterologously express diterpene synthases,[86,89,94,109,110] the enzymes that cyclize the 20-carbon precursor geranylgeranyl pyrophosphate (GGPP) into diterpene products.

Results and Discussion

To examine the utility of *Saccharomyces cerevisiae* for diterpene production, two genes from the *Arabidopsis thaliana* gibberellin biosynthetic pathway were expressed in a metabolically engineered yeast strain EHY18. EHY18 was previously developed in our lab to increase GGPP accumulation in vivo.[35,111] Yeast more readily converts the C-15

farnesyl pyrophosphate (FPP) precursor to squalene for use in the sterol biosynthetic pathway, resulting in limited production of GGPP. To overcome this, EHY18 was constructed by overexpressing the yeast GGPP synthase BTS1 in a yeast system containing an overexpressed truncated HMG-CoA reductase (trHMG1) in a *upc2-1* background (Figure 2.1).

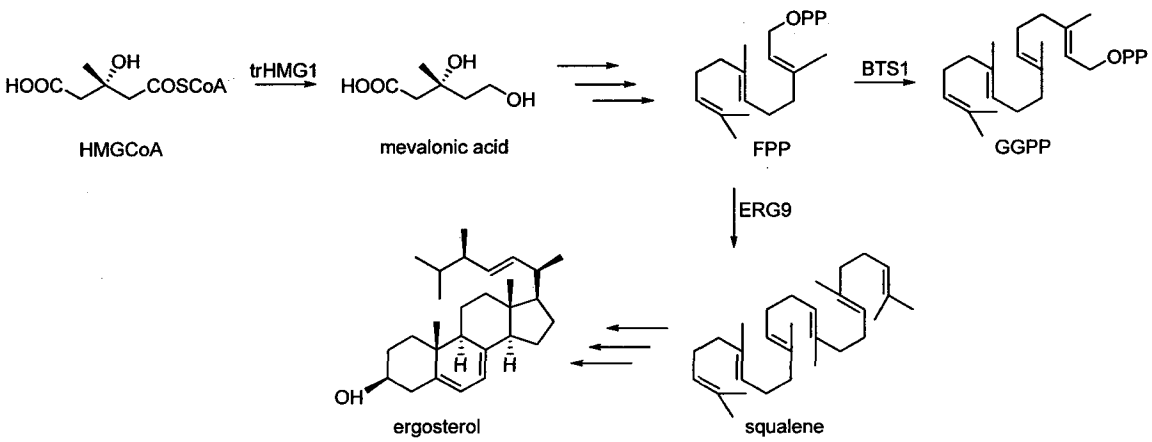

Figure 2.1. Modified sterol biosynthetic pathway in the EHY18 strain. trHMG1 and BTS1 were overexpressed to accumulate GGPP.

In gibberellin biosynthesis, two diterpene synthases are utilized in combination with a series of P_{450}-dependent oxidases to yield the final gibberellin products.[112] In *Arabidopsis*, GGPP is cyclized to *ent*-copalyl pyrophosphate by *ent*-copalyl pyrophosphate synthase (GA1),[113] and *ent*-kaurene synthase (GA2) further cyclizes *ent*-copalyl pyrophosphate to *ent*-kaurene (Figure 2.2).[114] Two yeast strains were constructed to produce the gibberellin precursors. First, the *Arabidopsis GA1* gene was expressed in EHY18 to accumulate *ent*-copalyl pyrophosphate, yielding the CPY1 strain (EHY18[pCVP2.1]). The CPY2 strain (CPY1[pCVP6.1]) contained both *Arabidopsis* genes *GA1* and *GA2* to produce *ent*-kaurene.

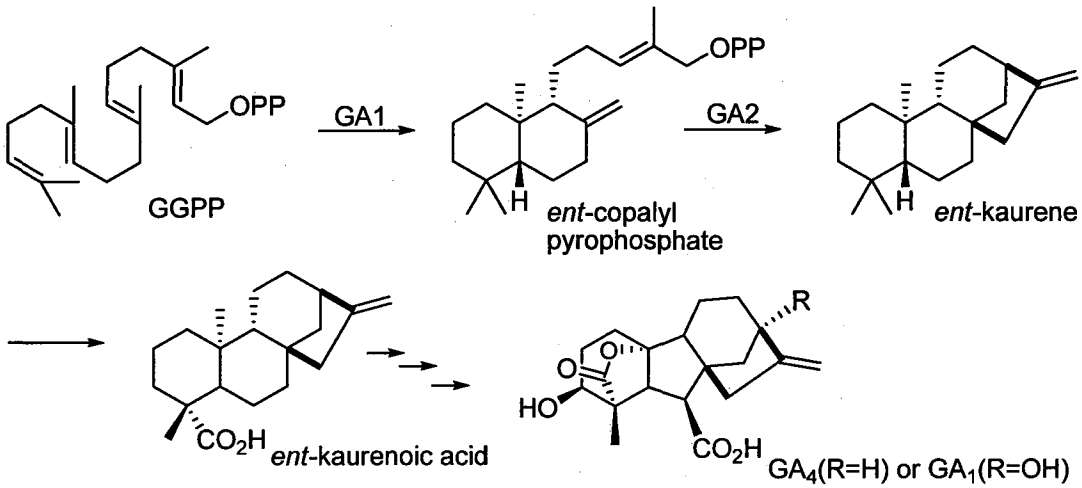

Figure 2.2. Gibberellin biosynthetic pathway. GGPP is cyclized to *ent*-copalyl pyrophosphate by GA1. Further cyclization is catalyzed by *ent*-kaurene synthase (GA2).

GC-MS analysis of CPY1 crude non-saponifiable lipids (NSL) revealed four main peaks with *m/z* ratios of 290, the mass expected for diterpene alcohols (Figure 2.3). Two peaks had mass spectra resembling cyclic diterpene alcohol products (See Appendix D for mass spectral data). ^{1}H NMR analysis confirmed the presence of *ent*-copalol,[115] a hydrolyzed version of *ent*-copalyl pyrophosphate, as the major cyclic product (Figure 2.4). The second peak was determined through ^{1}H NMR to actually be two cyclic products, *ent*-manool and *ent*-13-epimanool.[116] These two compounds co-eluted on GC-MS and could not be separated. No other cyclic products were detected within a 1% detection limit.

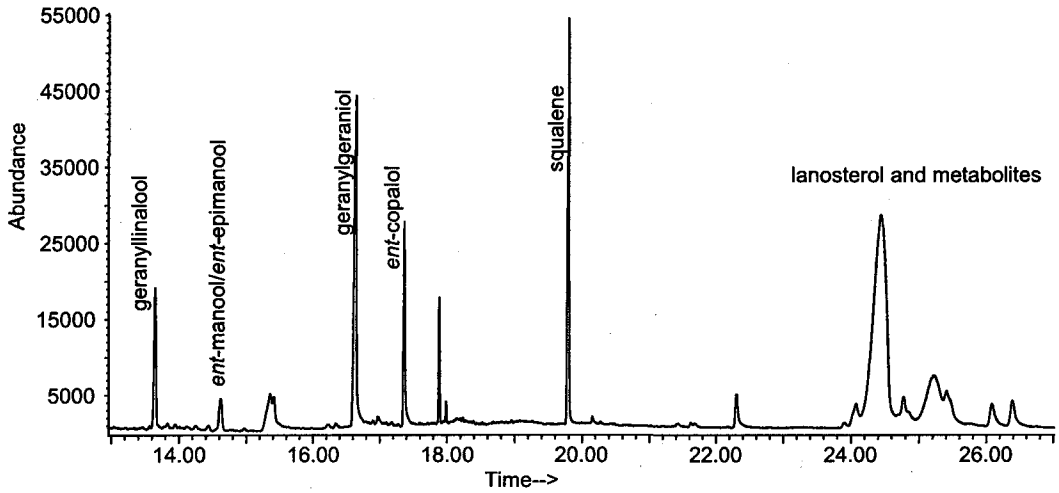

Figure 2.3. Total ion chromatogram of CPY1 crude NSL.

ent-Copalol, *ent*-manool, and *ent*-13-epimanool were most likely derived from hydrolysis of *ent*-copalyl pyrophosphate. Formation of *ent*-copalyl pyrophosphate begins with protonation of the GGPP terminal olefin and cyclization to form the bicyclic carbocation (Figure 2.4). Deprotonation at C-17 yields the exocyclic alkene and the final *ent*-copalyl pyrophosphate product. Once in an aqueous environment, the pyrophosphate moiety readily hydrolyzed to the three products. ^{1}H NMR analysis showed *ent*-manool and *ent*-13-epimanool in equal amounts (based on peak heights), suggesting non-enzymatic hydrolysis. The tertiary alcohols were found in much smaller amounts than the primary alcohol, suggesting an S_N2 mechanism was favored over S_N2' or S_N1 mechanisms to the tertiary alcohols.

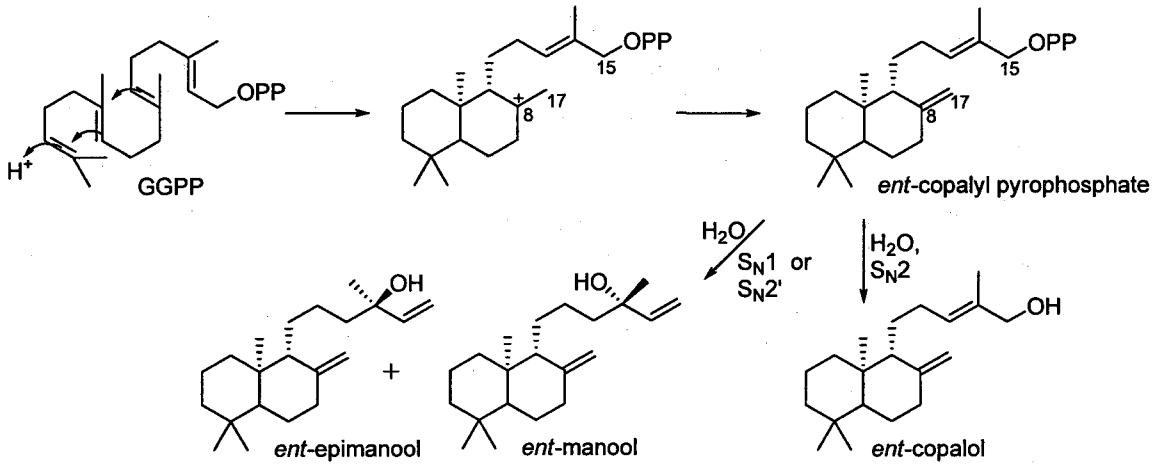

Figure 2.4. Proposed mechanism to *ent*-copalol, *ent*-manool, and *ent*-13-epimanool.

Two additional compounds were identified as geranylgeraniol and geranyllinalool (Figure 2.5), based on retention times and mass spectra identical to authentic standards (Sigma). Geranylgeraniol and geranyllinalool were also detected in the control strain EHY18[pRS426Gal], verifying that they are formed from hydrolysis of GGPP (Figure 2.4). EHY18 was previously reported to accumulate the primary alcohol geranylgeraniol,[35] and DeJong et al. reported geranyllinalool as a second hydrolysis product in their diterpene expression work in yeast,[89] verifying our results.

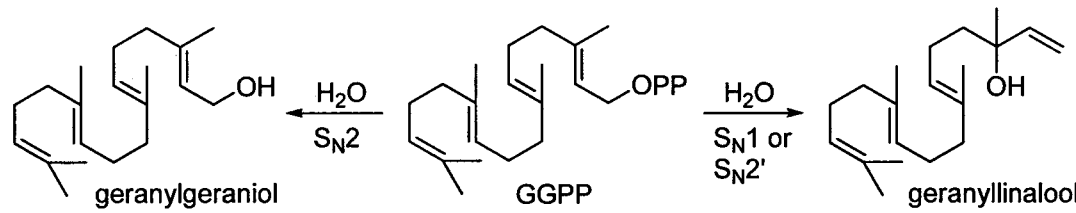

Figure 2.5. GGPP hydrolysis to geranylgeraniol and geranyllinalool.

The strain co-expressing both GA1 and GA2 was also analyzed, and five major diterpene peaks were revealed in the GC-MS analysis of the non-saponifiable lipids

(Figure 2.6). Four of these peaks corresponded to the diterpene alcohols found in the GA1 strain, identified as geranylgeraniol, geranyllinalool, *ent*-copalol, and *ent*-manool/*ent*-13-epimanool. The fifth peak had a mass spectrum resembling that of *ent*-kaurene, with an *m/z* ratio of 272, corresponding to that of a diterpene hydrocarbon. No other cyclic products were detected within a 1% detection limit.

As no reliable ^{1}H NMR data were found, Dr. William K. Wilson performed quantum mechanical calculations to predict the ^{1}H NMR chemical shifts, and the observed chemical shifts matched closely. Extensive 2D NMR analysis confirmed the structure as *ent*-kaurene (See Appendix D for chemical shift assignments).

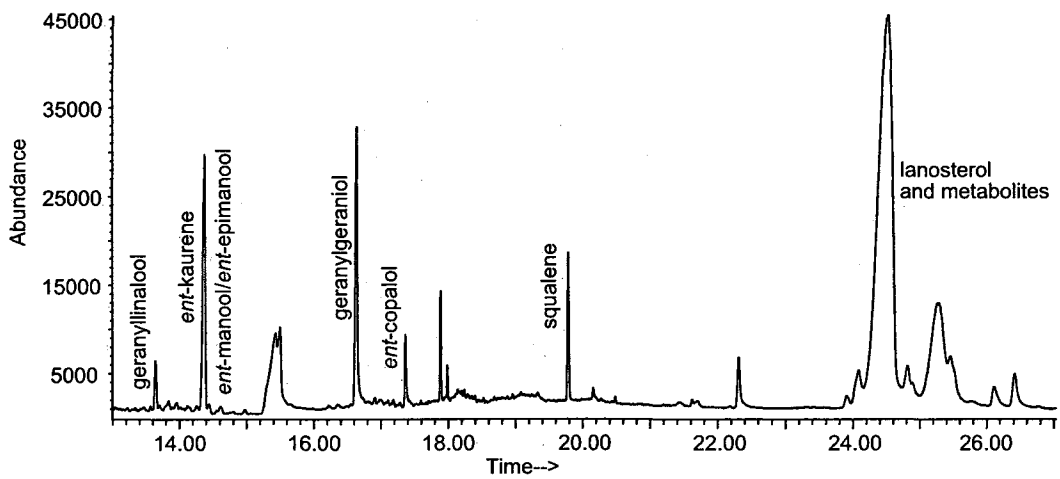

Figure 2.6. Total ion chromatogram of CPY2 crude NSL.

ent-Kaurene synthases employ a different cyclization mechanism than *ent*-copalyl pyrophosphate synthases (Figure 2.7). Whereas *ent*-copalyl pyrophosphate synthesis begins with protonation of the terminal double bond in GGPP, *ent*-kaurene biosynthesis starts with formation of a carbocation as a result of loss of the pyrophosphate moiety. The third ring forms, followed by deprotonation at C-17 to yield the final tetracyclic product.

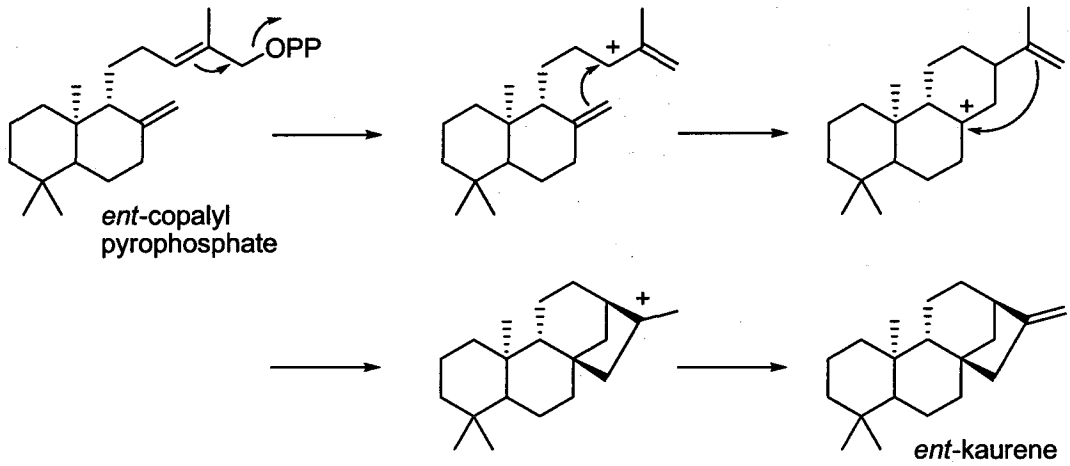

Figure 2.7. Proposed mechanism for *ent*-kaurene cyclization.

In addition to the gibberellin precursors identified, sterol metabolites were also recovered from the cultures. The EHY18 strain contains an intact sterol biosynthetic pathway, therefore the sterol precursors squalene and lanosterol, the membrane sterol ergosterol, and several lanosterol metabolites were detected in these yeast strains. That these were found at such high levels, and that not just ergosterol was isolated, suggested that the volume of substrate was too great for complete conversion to ergosterol or GGPP. It also indicated that a majority of the FPP was still being funneled into the sterol biosynthetic pathway and not towards GGPP.

Once characterization was complete, the three strains were quantitated by growing triplicate cultures to saturation in inducing media. The cultures were harvested, and the non-saponifiable lipids (NSL) were quantitated by GC-MS and GC-FID. The peak areas of geranylgeraniol, geranyllinalool, *ent*-copalol, *ent*-manool/*ent*-epimanool, and *ent*-kaurene were compared to that of a known amount of the internal standard epicoprostanol (Table 2.1).

Table 2.1. GC-FID quantitation data of EHY18[pRS426Gal], CPY1, and CPY2. All values are reported in mg/L, with peak areas compared to that of epicoprostanol.

Strain	Geranyl linalool	Geranyl geraniol	*ent*-Manool/*ent*-Epimanool	*ent*-Copalol	*ent*-Kaurene
EHY18 [pRS426Gal]	0.23 ± 0.046	1.68 ± 0.09	N/A	N/A	N/A
CPY1	0.22 ± 0.074	1.03 ± 0.17	0.14 ± 0.02	0.53 ± 0.04	N/A
CPY2	0.13 ± 0.013	0.46 ± 0.05	0.05 ± 0.001	0.19 ± 0.03	0.97 ± 0.08

The CPY1 strain produced approximately 0.53 mg/L of *ent*-copalol, along with 0.14 mg/L of *ent*-manool and *ent*-epimanool. CPY2, on the other hand, produced less of these bicyclic diterpene alcohols, with a majority of the *ent*-copalyl pyrophosphate being cyclized by GA2 to form *ent*-kaurene. The *ent*-kaurene levels were the highest cyclic product yields of the two strains, with nearly 1 mg/L accumulated. No cyclic diterpenes were recovered from the media (data not shown).

The control strain EHY18[pRS426Gal] accumulated 1.9 mg/L of the GGPP hydrolysis products geranylgeraniol and geranyllinalool. CPY1 accumulated approximately 1.2 mg/L of linear diterpene alcohols, and CPY2 contained the smallest amount of geranylgeraniol and geranyllinalool, with only 0.57 mg/L detected. All three strains produced more geranylgeraniol than geranyllinalool, suggesting the S_N2 mechanism to the primary alcohol occurred more readily than the S_N2' or S_N1 mechanism to the tertiary alcohol.

GA2 appears to be a more efficient enzyme than GA1, because the co-expressed pair accumulates few GA1 products and GGPP hydrolysis products. GA2 was able to cyclize a majority of *ent*-copalyl pyrophosphate, with only one-third the amount of *ent*-copalol, *ent*-manool, and *ent*-epimanool recovered. CPY2 also accumulated the lowest

levels of geranylgeraniol and geranyllinalool, suggesting cyclization occurred faster than hydrolysis. This is in sharp contrast to the GA1 strain, in which the GGPP hydrolysis products were more abundant than the GA1 cyclization products.

Overall, the three strains accumulated nearly 2 mg/L of total diterpenes, including both linear GGPP hydrolysis products and cyclization products, suggesting the same amount of carbon was available for the diterpene pathway in each strain.

This work demonstrates the utility of *S. cerevisiae* for diterpene production. A yeast strain with five modifications is still viable and producing milligram-scale amounts of diterpene compounds. This technology can also be applied to other commercially important diterpenes, sesquiterpenes, and triterpenes. The use of modified yeast strains as hosts for simultaneous expression of multiple genes, including terpene synthases and P_{450} oxidases, can lead to the development of larger libraries of terpenoid products with pharmaceutical or commercial potential.

Furthermore, this work demonstrates the level of competing processes occurring in the yeast strains. Geranylgeraniol and geranyllinalool were major diterpene products, resulting from hydrolysis of GGPP. These competing pathways divert GGPP away from the diterpene synthases, resulting in less cyclized products. If a yeast GGPP phosphatase were identified, deleting it could decrease GGPP hydrolysis,[74-76] increasing the amount of GGPP available for diterpene synthases.

Another source of competition is from the sterol pathway itself. Squalene, lanosterol, and other ergosterol intermediates were found in substantial amounts in the extracts, suggesting most of the FPP was utilized for sterol biosynthesis. FPP is converted to the sterol precursor squalene by squalene synthase, so deletion of squalene synthase

would remove a major competing pathway. Coupled with overexpression of GGPP synthase, this modification would likely increase GGPP accumulation, and thus diterpene cyclization products, in vivo. Deletion or repression of squalene synthase has greatly improved sesquiterpene cyclization product yields,[84,85,117] and it seems likely that this manipulation could aid in diterpene production as well.

Signaling sequences are another limitation affecting diterpene production in vivo. Plants produce diterpenes within plastids[118] and synthases typically contain signaling sequences targeting the proteins to those organelles. These signaling sequences are then removed in the plant because they compromise catalysis. The absence of this machinery in yeast causes lower yields in vivo.[92,93] The next chapter describes a method that increases diterpene synthase expression through inclusion of a chloroplast processing enzyme in the yeast strains.

Experimental Procedures

Subcloning of *GA1* and *GA2*. The plasmid pGA1-29 containing the *Arabidopsis* GA1 gene in pBluescript was from Tai-ping Sun (Duke University) and Yuji Kamiya (RIKEN).[113] Marligen ion-exchange columns were used to purify the plasmid DNA, following the manufacturer's protocol. The plasmid was digested with *Sal* I and *Not* I, gel-purified, and ligated into the yeast expression vector pRS426GalR that had been digested with the same enzymes. The resulting plasmid was named pCVP2.1.

The *Arabidopsis* GA2 plasmid was previously cloned into pRS305biGal by former lab member Kelly McCann. The yeast expression vector pRS313biGal and the GA2 plasmid DNA were digested with *Xba* I and *BamH* I and the fragments were gel-

purified and ligated, yielding the plasmid pCVP6.1. The pRS313biGal vector contains a HIS3 marker that allowed for selection during yeast transformations.

Construction of yeast strains. EHY18 (*MAT***a** *pGAL1-BTS1::his pGAL1-trHMG1::LEU2 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2* Gal^+)[35] was transformed with the pCVP2.1 plasmid using the lithium acetate method,[119] with selection on synthetic complete medium lacking uracil. This yeast strain was named CPY1 (EHY18[pCVP2.1]). The CPY1 strain was transformed with the pCVP6.1 plasmid, and the resulting transformants were selected on media lacking uracil and histidine. This strain, CPY1pCVP6.1], was named CPY2. A control strain was also prepared, in which the yeast expression vector pRS426Gal[120] was transformed into EHY18, with selection on synthetic complete media lacking uracil.

Large-scale analysis of GA1 strain. A 1-L culture of CPY1 was grown in inducing medium and harvested at saturation. The cell pellet was saponified, extracted, and analyzed by GC-MS. The crude NSL was then purified by column chromatography using an 80:20 mixture of CH_2Cl_2/hexane. Fractions were combined and analyzed by GC-MS and ^1H NMR. ^1H NMR analysis confirmed the identities of the *ent*-copalol structure by comparison with ^1H NMR literature data (Table 2.2).[115] Additionally, the *ent*-manool (Table 2.4) and *ent*-13-epimanool (Table 2.5) ^1H NMR chemical shifts also matched those reported in literature.[116]

Table 2.2. Literature values and observed ^{1}H NMR chemical shifts for *ent*-copalol. All values are reported in ppm.

Atom #	Literature value	Observed value	Difference
14-a	5.38, tq, J = 7.0, 1.3 Hz	5.392, tq, *J* = 1.3, 7.0 Hz	0.012
15a/b	4.15, d, J = 6.9 Hz	4.155, d, *J* = 6.9 Hz	0.005
16	1.65	1.66, br s	0.01
17-a	4.50, br d, J = 1.3 Hz	4.512, q, *J* = 1.6, 3.2 Hz	0.012
17-b	4.82, app q, J = 1.5 Hz	4.827, q, *J* = 1.7, 3.3 Hz	0.007
18-Me	0.86, s	0.870, s	0.01
19-Me	0.79, s	0.801, s	0.011
20-Me	0.67, s	0.679, s	0.009

Table 2.3. Literature values and observed ^{1}H NMR chemical shifts for *ent*-manool. All values are reported in ppm.

Atom #	Literature value	Observed value	Difference
14	5.90, dd, *J* = 10.7, 17.4 Hz	5.910, dd, *J* = 10.8, 17.4 Hz	0.01
15-a	5.03, dd, *J* = 1.3, 10.7 Hz	5.058, dd, *J* = 1.3, 10.8 Hz	0.028
15-b	5.19, dd, *J* = 1.3, 17.4 Hz	5.208, dd, *J* = 1.3, 17.4 Hz	0.018
16	1.26, s	1.271, s	0.011
17-a	4.46, d, *J* = 1.4 Hz	4.475, q, *J* = 1.6, 3.1 Hz	0.015
17-b	4.80, d, *J* = 1.4 Hz	4.804, q, *J* = 1.6, 3.1 Hz	0.004
18-Me	0.85, s	0.866, s	0.016
19-Me	0.78, s	0.797, s	0.017
20-Me	0.66, s	0.671, s	0.011

Table 2.4. Literature values and observed ^{1}H NMR chemical shifts for *ent*-13-epimanool. All values are reported in ppm.

Atom #	Literature value	Observed value	Difference
14	5.90, dd, *J* = 10.7, 17.4 Hz	5.912, dd, *J* = 10.8, 17.4 Hz	0.012
15-a	5.03, dd, *J* = 1.3, 10.7 Hz	5.048, dd, *J* = 1.3, 10.8 Hz	0.018
15-b	5.19, dd, *J* = 1.3, 17.4 Hz	5.201, *J* = 1.3, 17.4 Hz	0.011
16	1.26, s	1.273, s	0.013
17-a	4.50, d, *J* = 1.5 Hz	4.512, q, *J* = 1.6, 3.1 Hz	0.012
17-b	4.80, d, *J* = 1.5 Hz	4.815, q, *J* = 1.5, 3.1 Hz	0.015
18-Me	0.85, s	0.866, s	0.016
19-Me	0.78, s	0.797, s	0.017
20-Me	0.66, s	0.677, s	0.017

Large-scale analysis of GA2 strain. A 1-L culture of CPY2 was grown to saturation in galactose and harvested by centrifugation. The cell pellet was saponified and

extracted, and the crude NSL was analyzed by GC-MS. The NSL was then purified by column chromatography (6 g silica gel 60) with a gradient of hexane/ether, starting with 100% hexane. The pure hexane fractions were further analyzed by ^{1}H and ^{13}C NMR, and COSY, DEPT, and HMBC 2D NMR experiments were also undertaken. All spectra were collected using a Bruker 500 MHz spectrometer. Chemical shifts were compared to values predicted by quantum mechanical calculations (Table 2.5) and ^{13}C literature values (Table 2.6 and Appendix C).[121]

Table 2.5. Predicted values and observed ^{1}H NMR chemical shifts for *ent*-kaurene. Asterisks (*) or daggers (†) indicate chemical shifts could be interchanged. All values are reported in ppm.

Atom #	Predicted value	Observed value	Difference
1-a	0.73	0.738, tdd, J = 13, 4.0, 0.9 Hz	0.008
1-b		1.806, dtd, J = 12.8, 3.5, 1.6 Hz	
2-a		1.393, m	
2-b		1.610, m	
3-a		1.122, td, J = 13.6, 4.6 Hz	
3-b		1.373, m	
5	0.82	0.787, dd, J = 12.0, 2.0 Hz	-0.033
6-a		1.547, m	
6-b		1.323, m	
7-a		1.516, m	
7-b		1.493, m	
9		1.058, m	
11-a		1.586, m	
11-b		1.586, m	
12-a		1.647, m	
12-b		1.471, m	
13	2.63	2.631 br t, J = 5.6 Hz	0.001
14-a		1.994, dd, J = 11.4, 2.5 Hz	
14-b		1.094, dddd, J = 11.4, 2.4, 2.4, 2.0 Hz	
15-a		2.064†, ddd, J = 3.0, 2.3, 0.9 Hz	
15-b		2.042†, ddd, J = 3.0, 2.3, 1.0 Hz	
17-a	4.69	4.729*	0.039
17-b	4.78	4.789*	0.009
18-Me	0.83	0.853, s	0.023
19-Me	0.81	0.810, s	0
20-Me	1.01	1.020, s	0.010

Table 2.6. Comparison of literature and observed ^{13}C NMR chemical shifts for *ent*-kaurene. All values are reported in ppm.

Atom #	Literature value	Observed value	Difference
1	41.3	40.43	-0.87
2	18.7	18.66	-0.04
3	42.0	42.08	0.08
4	33.3	33.27	-0.03
5	56.1	56.27	0.17
6	20.3	20.25	-0.05
7	40.4	41.24	0.84
8	44.2	44.23	0.03
9	56.1	56.07	-0.03
10	39.3	39.33	0.03
11	18.1	18.15	0.05
12	33.3	33.31	0.01
13	44.2	44.05	-0.15
14	39.9	39.84	-0.06
15	49.2	49.22	0.02
16	156.0	156.20	0.2
17	102.8	102.79	-0.01
18	33.7	33.66	-0.04
19	21.7	21.64	-0.06
20	17.6	17.59	-0.01

Quantitation of yeast strains. Triplicate 100-mL galactose cultures were grown to saturation and harvested by centrifugation at 3000 rpm for 10 min. The cell pellets were saponified in 10% KOH in 80% EtOH. To this were added the antioxidant BHT (2 mg/g cell pellet) and 100 μg of epicoprostanol as internal standard prior to 2 h incubation at 70 °C. Once the mixtures were cooled to room temperature, 5 mL distilled water was added, and the mixtures were extracted with 3 × 15 mL hexanes. The solvent was removed in vacuo, and the residues were transferred to GC vials with CH_2Cl_2 and analyzed by GC-MS and GC-FID.

Chapter 3

The production of mature plastid proteins in yeast: co-expressing a chloroplast processing enzyme facilitates metabolic engineering and enzyme investigation in diterpenes

Introduction

Biosynthesis of diterpenes in plants occurs in plastids,[122-124] and therefore diterpene synthases usually contain N-terminal transit peptide signal sequences directing them to the plastids.[90] Once in the organelle, the transit sequences are cleaved to form the mature, functional diterpene synthases. The initial full-length protein containing the transit peptide sequence expressed in *E. coli* has poor enzymatic activity, whereas removal of the signal sequence has been demonstrated to increase enzymatic activity and aid in enzyme characterization.[92-94]

Pseudomature proteins are heterologously expressed proteins lacking the transit peptide. These enzymes are generated by making a subclone without the signal sequence, but with an artificial start site introduced at the beginning of expected mature protein. Pseudomature proteins differ from conventional mature proteins because they retain this adventitious methionine, and pseudomature diterpene synthases often have better activity than their full-length counterparts when expressed in *E. coli*.[87,93,95] However, the exact location of the cleavage site is not readily predictable, therefore extensive empirical experiments are often necessary to determine the best site for N-terminal truncation.[93,95,96,125] Moreover, one cannot be certain that the introduced methionine in the pseudomature protein does not alter catalysis in some way. However, plants encode

proteins known as chloroplast processing enzymes (CPE), which cleave transit peptides once proteins enter the plastid.[91] Described herein are experiments establishing that co-expression of a chloroplast processing enzyme with a diterpene synthase promotes truncation of the signaling sequence within the heterologous host, removing the necessity for manual truncation through PCR amplification. This approach generates the mature diterpene synthase, increasing diterpene synthase biosynthetic activity in our yeast system.

Results and Discussion

The *Arabidopsis* chloroplast processing enzyme (CPE, *At5g42390*) was PCR-amplified from a cDNA and co-expressed in EHY18[35,111] containing either the *Abies grandis* (grand fir) abietadiene synthase (*Ag*AS)[126] or the *Arabidopsis ent*-copalyl pyrophosphate synthase (*At*GA1)[113] and *ent*-kaurene synthase (*At*GA2)[114] genes (Table 3.1). Control strains expressing only diterpene synthases were grown in parallel, and the diterpene products accumulating in the cell pellets and media were analyzed. Triplicate cultures were grown and harvested to provide more accurate data.

Table 3.1. Summary of yeast strains with and without CPE. All yeast strains contain *MAT***a** *pGAL1-BTS1::hisG pGAL1-trHMG1::LEU2 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2* Gal^+.

Yeast strain	Characteristics
EHY18[pEH9.0]	*AgAS::URA3*
EHY18[pEH9.0][pAMB5.3]	*AgAS::URA3 AtCPE::HIS3*
EHY18[pCVP26.1]	*AtGA1+AtGA2::URA3*
EHY18[pCVP26.1][pAMB5.3]	*AtGA1+AtGA2::URA3 AtCPE::HIS3*

The cell pellets from EHY18[pEH9.0] and EHY18[pEH9.0][pAMB5.3] were saponified, and the non-saponifiable lipids (NSL) contained the majority of the diterpene products (Table 3.2). The NSL were thus used for quantitation, as insignificant levels of diterpene products were isolated from the media (<1% of the total). GC-MS analysis of the NSL showed that the cultures produced abietadiene as the major diterpene product. Abietadiene synthase was previously reported to catalyze cyclization of geranylgeranyl pyrophosphate (GGPP) to the intermediate (+)-copalyl pyrophosphate, which is then cyclized further to (-)-abieta-7(8),13(14)-diene (Figure 3.1).[126]

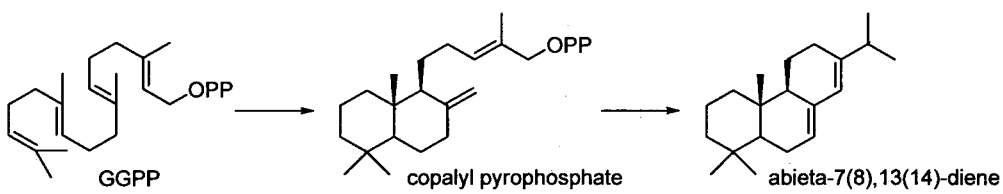

Figure 3.1. Cyclization of GGPP to abietadiene by abietadiene synthase.

Table 3.2. GC-FID quantitation of EHY18[pEH9.0] NSL with and without CPE co-expression. Peak areas were compared to that of the internal standard epicoprostanol.

Strain	Abietadiene (mg/L)	Geranylgeraniol (mg/L)	Squalene (mg/L)
EHY18[pEH9.0]	5.91 ± 0.55	N. D.	14.18 ± 2.16
EHY18[pEH9.0][pAMB5.3]	9.11 ± 0.58	N. D.	8.29 ± 1.02

The amount of abietadiene isolated from the CPE-expressing strain increased 1.5 times in comparison to the strain without the CPE. Both strains accumulated a significant amount of abietadiene, with nearly 6 mg/L in the strain without the CPE and over 9 mg/L in the strain with the CPE. Unlike the diterpene systems described in Chapter 2, both the control strain and the strain co-expressing the CPE produced abietadiene but did not

accumulate substantial levels of the GGPP hydrolysis products geranylgeraniol or geranyllinalool.

Because the yeast strains have intact sterol biosynthetic pathways, squalene was also isolated but at much higher levels than the diterpene products. In the strain without the CPE, approximately 14 mg/L of squalene was accumulated, suggesting that the amount of carbon flux was too great for the enzymes in the sterol pathway to process it efficiently. The strain with the co-expressed CPE, however, demonstrated a large decrease in squalene accumulation, with only 8.3 mg/L recovered. This further supports the observation that CPE co-expression increases abietadiene synthase activity in the yeast.

Usually, most FPP funnels through the sterol biosynthetic pathway instead of being used for GGPP production. Overexpression of GGPP synthase shunted more FPP to GGPP biosynthesis and therefore provided more substrate for diterpene synthesis, decreasing the amount of carbon routed towards sterol biosynthesis.[35]

More dramatic changes were observed in the strains co-expressing GA1 and GA2 with the CPE (Table 3.3). As seen in the strains containing abietadiene synthase, almost all of the *ent*-kaurene was isolated from the cell pellets, with less than 1% of the *ent*-kaurene found in the media. From these results, the amount of *ent*-kaurene accumulating in the CPE co-expression strain was 10 times higher than that observed in the strain without the CPE. The parent strain accumulated approximately 0.25 mg/L *ent*-kaurene, while the strain with the CPE accumulated approximately 2.5 mg/L. This is the most *ent*-kaurene observed in any strain to date.

Table 3.3. GC-FID quantitation of EHY18[pCVP26.1] NSL with and without CPE co-expression. Peak areas were compared to that of the internal standard epicoprostanol.

Strain	*ent*-Kaurene (mg/L)	Geranylgeraniol (mg/L)	Squalene (mg/L)
EHY18[pCVP26.1]	0.25 ± 0.08	3.54 ± 1.26	8.18 ± 1.85
EHY18[pCVP26.1][pAMB5.3]	2.47 ± 1.34	2.46 ± 0.63	6.00 ± 0.73

The CPE co-expression greatly improved GA1 and GA2 activity, as shown by the substantial difference in *ent*-kaurene accumulation. No significant amounts of the *ent*-kaurene intermediates *ent*-copalol or *ent*-manool/*ent*-13-epimanool were observed (<1% of *ent*-kaurene), suggesting that GA2 enzymatic activity greatly increased and more efficiently cyclized *ent*-copalyl pyrophosphate to *ent*-kaurene.

In addition to *ent*-kaurene, a considerable amount of geranylgeraniol was also isolated, with approximately 3.0 mg/L accumulating in both EHY18[pCVP26.1] and the CPE-expressing strain. The large amount of geranylgeraniol isolated suggests that the GA1 and GA2 enzymes were not as efficient as the abietadiene synthase. As observed in the abietadiene strains, the amount of squalene accumulation was substantial, with over 8 mg/L of squalene isolated from EHY18[pCVP26.1]. The amount of squalene decreased significantly to 6 mg/L when the CPE was expressed. *ent*-Kaurene production increased at the same rate that squalene accumulation decreased, again consistent with CPE increasing GA1 and GA2 activity.

Co-expressing *Arabidopsis* CPE improved abietadiene production somewhat less than it did in the *ent*-kaurene-producing strains, consistent with better cleavage when CPE and cyclase both come from *Arabidopsis* than when they are from different organisms. This difference may reflect the evolutionary distance between *A. grandis* and

Arabidopsis. The CPE substrate recognition sites may not be similar enough to efficiently cleave the transit peptide from the other organism.

This work is the first example of improving the function of a plastid enzyme by co-expression with CPE. Some previous efforts generated pseudomature proteins by initiating transcription with a false start site after the signal sequence. However, this method has severe limitations. Transit peptides vary in length, from short sequences to up to 140 amino acids cleaved in some cases,[91] and variation in sequences make it difficult to predict the exact cleavage site.[109,127] Transit peptides lack defined sites that clearly indicate the precise location of the cleavage site.[128] The best existing methods to locate signal sequence boundaries rely on broad trends. The sequences are typically comprised of 20-35% serine and threonine as well as numerous small hydrophobic residues such as alanine, valine, or glycine.[129] While large amounts of basic amino acids, such as lysine or arginine, are not found, the peptides normally have an overall positive charge, with almost no acidic amino acid residues.[129]

The *A. grandis* abietadiene synthase and *Arabidopsis* GA1 and GA2 protein sequences, together with sequences from other characterized diterpene synthases, were downloaded from the National Center for Biotechnology Information (NCBI) website and analyzed for these general traits (Figure 3.2). All of these enzymes have putative transit peptides containing the proposed characteristics. Their N-terminal regions contained high levels of serine, threonine, and small hydrophobic molecules, and few acidic residues were observed. Somewhere between 75 and 140 residues, the density of acidic residues increased 5- to 10-fold, indicating the boundary to the mature protein had been crossed. However, no specific motifs marked the location of the native site.[86,93,96]

None of the peptides have obvious amino acid motifs in the putative transit peptide region, demonstrating the difficulty in determining the exact truncation site required.

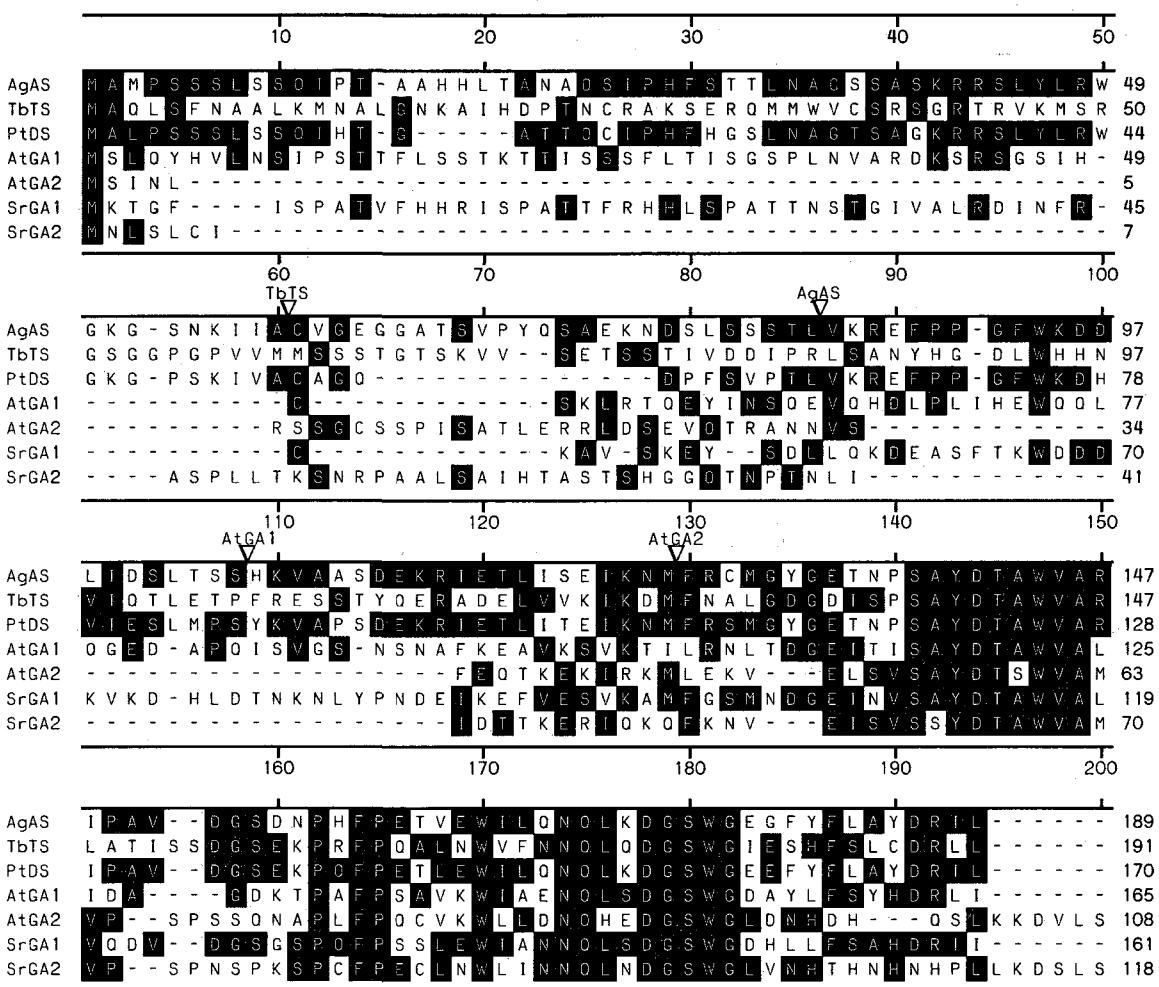

Figure 3.2. Protein alignment of the N-termini of seven diterpene synthases. AgAS is *A. grandis* abietadiene synthase, TbTS is *Taxus brevifolia* taxadiene synthase, PtDS is a *Pinus taeda* putative diterpene synthase, AtGA1 is *Arabidopsis* GA1, AtGA2 is *Arabidopsis* GA2, SrGA1 is *Stevia rebaudiana* copalyl pyrophosphate synthase, and SrGA2 is *S. rebaudiana ent*-kaurene synthase. Triangles denote the putative cleavage sites for AgAS, TbTS,[93] AtGA1, and AtGA2 . Areas of consensus sequences are shaded.

The signaling sequence in abietadiene synthase was initially suggested to comprise amino acids Met1-Ala110.[126] However, a series of truncations were

constructed, and truncation after Trp93 dramatically decreased enzymatic activity.[95] The conserved DDXXD motif necessary for catalysis[26] had been removed; the transit peptide was shorter than initially predicted. Peters et al. found the highest activity when the first 84 amino acids were cleaved.[95] Analysis of the first 84 amino acids of the NCBI abietadiene synthase sequence showed 29% serine and threonine, as well as 21% small hydrophobic residues. Only three acidic residues were found, consistent with the current best understanding of signal sequence characteristics.

More importantly, the CPE-cleaved abietadiene synthase working in vivo has a dramatically different product profile than the pseudomature proteins that have been developed to date. The native enzyme isolated from *A. grandis* was originally reported to generate abietadiene with high accuracy (over 90% abietadiene).[92,126] However, the pseudomature abietadiene synthases generate a diverse set of products in which abietadiene is a minor component.[95] In contrast, our metabolically engineered systems produce abietadiene at >95% of the product profile. Whether these differences are due to cleavage at a different site or the presence of the adventitious methionine remains to be determined.

Pseudomature derivatives of *Arabidopsis* GA1 showed similar trends to those seen in abietadiene synthase. The transit peptide in *Arabidopsis* GA1 was initially reported as approximately 50 amino acids.[113] However, Prisic and Peters later generated a series of pseudomature proteins at several sites along the GA1 protein and demonstrated that a 50-amino acid truncation yielded a very unstable protein.[96] Instead they reported an 84-amino acid cleavage that had better enzymatic activity than the immature enzyme.

This cleavage site was also at an exon boundary, which the authors argue makes this position a plausible location.[96]

However, the 84-amino acid sequence is more acidic, with 8% acidic residues and 7% basic residues. The region contained 26% serine and threonine, but these were present only in the first 60 amino acids. If this cleavage site is indeed correct, the guidelines for predicting signal sequences need revision. Without making every possible pseudomature protein along this region, the possibility exists that a shorter peptide could be deleted and retain good activity. A further possibility is that the adventitious methionine that initiates translation compromises enzymatic activity in the derivative from which 50 residues was deleted.

The transit peptide in *Arabidopsis* GA2 was also tentatively identified as 44 amino acids.[114] This region of GA2 had 25% serine or threonine and 18% basic residues. Notably, the sequence also contained a higher level of acidic residues, with 11% observed. No pseudomature proteins have been described for GA2 to date.

Because plastid proteins contain transit peptides, each plant must encode a CPE or a similar protein that cleaves the transit peptides within the plant. Chloroplast processing enzymes have been identified in many plants, including *Pisum sativa* and *Arabidopsis thaliana*.[91] These enzymes are metalloendopeptidases, which typically have rather broad substrate specificities. CPE was previously reported to process proteins targeted to many different locations within the chloroplast and to cleave proteins that serve roles in varied biosynthetic pathways.[91] Typically, the transit peptide is cleaved in a single proteolytic step, resulting in a mature protein, and the cleaved transit peptide is further degraded within the organelle.[130]

As a plastid protein itself, CPE is translated as a preprotein with a transit peptide that is cleaved once entering the plastid.[91] Analysis of the CPE protein sequence uncovered a putative transit peptide rich in serine and threonine residues and small hydrophobic amino acids, consistent with reported characteristics. The CPE preprotein has little activity in *E. coli*,[91] and its activity in yeast shows that the CPE apparently cleaves its own sequence as well to yield the mature protein, as no other chloroplast enzymes are present in our yeast strains. This is supported by Richter and Lamppa, who reported that the *P. sativa* CPE was capable of processing its own transit peptide in *E. coli*.[91]

The differences in product improvement between the abietadiene synthase from a gymnosperm and *ent*-kaurene synthase from *Arabidopsis* suggest that the *Arabidopsis* CPE may not be optimal for enzymes from other organisms. Subtle differences in the recognition sequences may be more efficiently cleaved by a CPE from the same organism or a close relative. The possibility that different CPEs may cleave different sites cannot be precluded. As more genomic sequence becomes available it will be straightforward to develop an array of yeast strains expressing CPEs from the major plant groups to optimize cleavage. However, we demonstrate that even without an organism-specific CPE, the *Arabidopsis* CPE improves diterpene yields significantly.

In addition to increasing diterpene production, this approach can be utilized for expression of any plastidial protein. For example, many monoterpenes are also produced in the plastids and a large number of monoterpene synthases encode a transit peptide. Co-expression of a chloroplast processing enzyme with a monoterpene synthase could potentially result in increase of monoterpene yields in vivo.

This work demonstrates a novel and facile approach to increasing diterpene production in yeast. Co-expression of a chloroplast processing enzyme removes the need for manual truncation of the diterpene synthase transit peptide and in many cases should provide a more active enzyme than pseudomature proteins. Previously, multiple truncations had to be examined to find the best protein in terms of solubility, purification, and enzymatic activity, but this approach allows in vivo cleavage to yield the mature protein. This method shows vast improvements compared to previously reported yields and demonstrates the great potential for this strategy.

Experimental Procedures

RT-PCR to obtain a cDNA. mRNA from 7-day *Arabidopsis* seedlings (Columbia) was a generous gift from Prof. Rebekah Rampey, a former graduate student from Prof. Bonnie Bartel's laboratory. Reverse-transcription PCR was performed following the manufacturer's protocol to obtain a cDNA (Ambion, Austin, TX).

PCR amplification of *CPE* from cDNA. The full-length *Arabidopsis CPE* gene sequence (At5g42390) was obtained from NCBI and used to develop primers for PCR amplification from the cDNA library. Because of the length of the *CPE* gene (3.8 kB), it was necessary to amplify the gene in two parts: from *Sal* I to *Eco*R I, and from *Eco*R I to *Not* I. For the first fragment, the PCR mixture was as follows: 1 μL 7-day seedling cDNA, 1 μL of the forward primer pCVP18-F1 5'-TAGTCGACAATTATGGCTTCATCG-3' (*Sal* I restriction site underlined), 1 μL of the reverse primer pCVP18-R2 5'-ACCGGAATTCCATGGCAATG-3' (*Eco*R I restriction

site underlined), 4 μL dNTPs, 5 μL High-Fidelity buffer, 0.2 μL Triple Master polymerase, and 38 μL mq$H_2$0. The back fragment contained the same mixture, except the primers used were pCVP18-F7 5'-CATTGCCATGGAATTCCGGTTTACT-3' (*Eco*R I restriction site underlined) and pCVP18-R1 5'-GCGGCCGCTCAGGTTGTTGGTCTTGT-3'(*Not* I restriction site underlined).

Construction of the full-length *CPE*-containing plasmid. The two *CPE* fragments, *Sal* I to *Eco*R I, and *Eco*R I to *Not* I, were ligated into TOPO-TA vector, and clones containing the inserts were sequenced. Large-scale cultures of clones with the correct sequences were grown and DNA isolated, yielding two plasmids pAMB3.0 and pAMB4.0, corresponding to the front and back pieces of the *CPE*, respectively.

The two plasmids pAMB3.0 and pAMB4.0 were digested with their corresponding restriction enzymes and ligated into pRS313Gal that had been digested with *Sal* I and *Not* I. The three-piece ligation was then transformed into *E. coli* and plated on LB + Amp. Clones were screened for the presence of the insert, and the correct clone was sequenced to ensure the full-length gene was obtained with the correct sequence. The final plasmid was named pAMB5.3.

Subcloning *GA1* and *GA2* into one plasmid. A plasmid containing both *Arabidopsis GA1* and *GA2* was constructed. First, *GA2* and a bi-directional galactose (BiGal) promoter were subcloned into pRS426 with *BamH* I and *Not* I to give the plasmid pCVP6.3. *GA1* was PCR-amplified to insert new restriction sites using the primers GA1F4 5'- TACCGCGGAATTATGTCTCTTCAGTATCATG-3' and GA1NotIR 5'-

GCGGCCGCCTAGACTTTTTGAAACAAG_-3'. The gel-purified PCR product was cloned into pCVP6.3 with *Sac* II and *Not* I, resulting in the final plasmid pCVP26.1.

Construction of yeast strains. Two control strains were first constructed. EHY18 was transformed with either the pCVP26.1 plasmid or pEH9.0, the *Abies grandis* abietadiene synthase in pRS426Gal (constructed by former lab member Dr. Elizabeth Hart). These two yeast strains, EHY18[pCVP26.1] and EHY18[pEH9.0], were selected on synthetic complete media lacking uracil and transformed with the pAMB5.3 plasmid. The final transformants were screened by growing on synthetic complete media lacking both uracil and histidine.

Identification and quantitation of products generated in yeast strains. All strains were grown to saturation in inducing medium and harvested by centrifugation. The cell pellets were saponified and extracted with 3 × 15 mL hexanes, with 100 μL of the internal standard epicoprostanol (2.5 mg in 1 mL ethanol) added prior to saponification. The NSL were partitioned between one-half volume water and 3 × 15 mL hexanes, and solvent was evaporated in vacuo and residues transferred to GC vials with CH_2Cl_2. Samples were analyzed on GC-FID and GC-MS.

Chapter 4

Manipulation of media pH to increase nerolidol or farnesol production

Introduction

Metabolic engineering has recently emerged as an alternative to obtaining terpenoids from natural sources and synthesis. *Saccharomyces cerevisiae* ia a particularly attractive host for heterologous expression of sesquiterpene biosynthetic genes because fungi biosynthesize the sesquiterpenoid precursor farnesyl pyrophosphate (FPP) as an intermediate in sterol biosynthesis (Figure 4.1). Several yeast strains have been developed to increase FPP accumulation, focusing on increasing the amount of substrate available through a combination of deleting or repressing competing pathways and increasing carbon flux through the pathway. For example, a solubilized form of the rate-limiting enzyme HMG-CoA reductase (trHMG1) has been overexpressed to increase the amount of carbon available for the pathway, and the competing enzyme squalene synthase (ERG9) has been repressed or deleted in an effort to increase FPP accumulation further.[82-85,117,131] However, FPP appears to hydrolyze readily when it accumulates in yeast. Strains with deleted or repressed ERG9 produce farnesol[41,132] (Figure 4.2) as a dominant product of FPP hydrolysis, and when sesquiterpene synthases (the enzymes cyclizing FPP to hydrocarbon products) are expressed in these strains, farnesol predominates over cyclized products.[131]

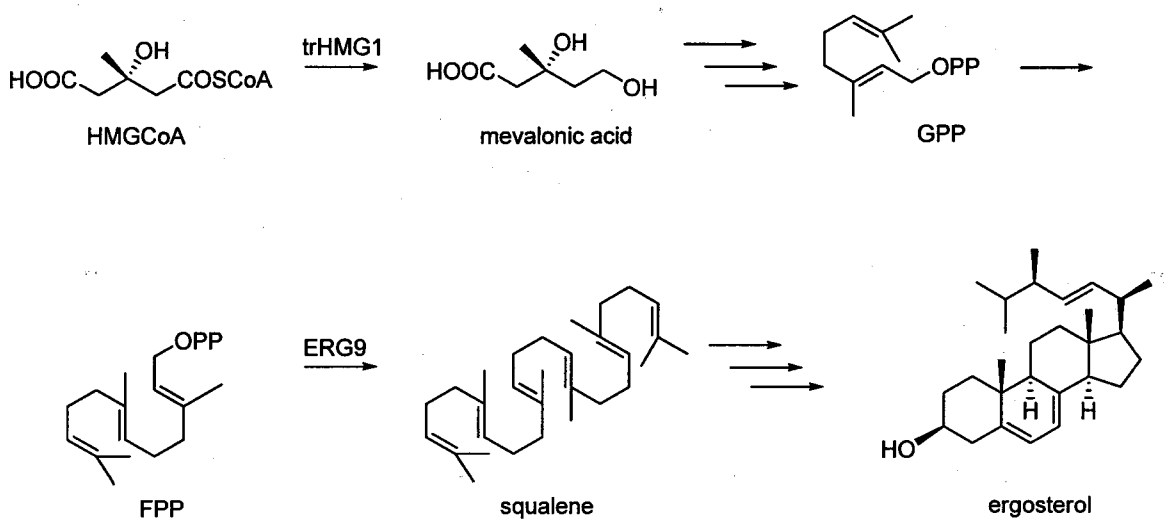

Figure 4.1. *S. cerevisiae* sterol biosynthetic pathway. ERG9 (squalene synthase) and trHMG1 (HMG-CoA reductase) are noted.

Results and Discussion

We constructed the yeast strain PMY1, an *erg9* mutant that overexpresses trHMG1.[82] The trHMG1 overexpression promotes FPP biosynthesis, which cannot be converted to squalene because of the *erg9* deletion, and we found that PMY1 accumulates two major FPP hydrolysis products, the primary alcohol farnesol and the tertiary alcohol nerolidol (Figure 4.2).

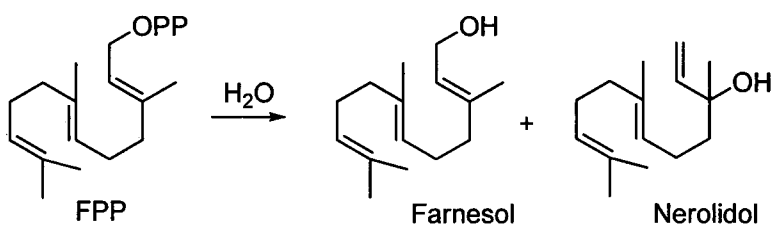

Figure 4.2. Hydrolysis of FPP forms the primary alcohol farnesol or the tertiary alcohol nerolidol.

S. cerevisiae encodes two phosphatases, DPP1[74] and LPP1,[75] which hydrolyze FPP and other prenyl pyrophosphates.[76] We speculated that one phosphatase might generate farnesol through an S_N2 mechanism, while the other could form nerolidol by S_N2' or S_N1 displacements. We investigated the impact of these phosphatases on the metabolically engineered profile through mutagenesis experiments. Each phosphatase was deleted by replacing the gene with a geneticin-resistant marker, *KanMX*, to give PMY2 (*dpp1Δ*) and BEJY14 (*lpp1Δ*). These were expected to substantially reduce the amounts of FPP hydrolysis products produced in these yeast strains, as DPP1 was previously implicated in decreasing farnesol production in engineered yeast.[131] The sesquiterpene content of media and cell pellets were assayed. FPP hydrolysis products were found not in the cell pellet, but were isolated in the media (Table 4.1).

Table 4.1. PMY1, PMY2, and BEJY14 media extracts were quantitated by GC-FID. Farnesol and nerolidol values are reported in mg/L. All strains were grown in YPGHE and possess *MAT***a** *erg9::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 his3-Δ200 ade2 GAL+*.

Strain	Genotype	Farnesol (mg/L)	Nerolidol (mg/L)
PMY1	trHMG1::LEU2	28.6 ± 1.35	3.10 ± 0.061
PMY2	*dpp1*Δ::KanMX trHMG1::LEU2	16.7 ± 1.23	0.943 ± 0.0772
BEJY14	*lpp1*Δ::KanMX trHMG1::LEU2	28.3 ± 3.09	1.18 ± 0.201

Although the phosphatase deletions had statistically significant effects on farnesol and nerolidol production, these changes were not dramatic enough to suggest that LPP1 and DPP1 are the primary modes of FPP hydrolysis. The total amounts of $C_{15}H_{26}O$ compounds were not impacted by the *lpp1Δ* deletion, but the *dpp1Δ* strain accumulated somewhat less of these compounds. Farnesol production was not significantly impacted by deleting LPP1 but was significantly reduced in the *dpp1*Δ strain. This led to the

conclusion that DPP1 played a role in farnesol production, but whether it directly mediated FPP hydrolysis to farnesol in vivo or altered sterol precursor biosynthesis and metabolism indirectly remains to be determined. One striking observation was that the major product did not vary between the three strains; farnesol production was still the major outcome. The farnesol to nerolidol ratio of PMY1 (9.22) was somewhat lower than that of PMY2 (17.7) and BEJY14 (24.0). Nerolidol production was decreased by deleting either LPP1 or DPP1. The consistency in nerolidol data between PMY2 and BEJY14, however, suggested this S_N2' or S_N1 displacement to the tertiary allylic alcohol occurred, in part, outside of the environment of LPP1 or DPP1.

Because no other phosphatase seemed an obvious candidate for the enzyme that hydrolyzes FPP to nerolidol, the possibility that these compounds are nonenzymatic products was investigated. The stereochemistry of nerolidol isolated and purified from PMY1[pRS316Gal] was established with ^{1}H NMR analysis using a lanthanide chiral shift reagent. Chiral shift reagent NMR analysis[133] has been well established and is often more reliable than polarimetry.[134] Whereas polarimetry is susceptible to changes in solvent, concentration, and trace impurities, chiral shift reagent NMR is a non-optical method to distinguish enantiomers based on differences in chemical shifts after addition of a chiral shift reagent. Additionally, very little material is needed for NMR analysis, and the use of a high-field magnet enables fine analysis with high precision. Comparison of peak areas or peak heights can then determine enantiomeric ratios. Linalool is a monoterpene homolog of nerolidol, and the lanthanide shift reagent Eu(hfc)$_3$ (europium tris[2-(heptafluropropylhydroxymethylene)-(+)-camphorate]) was reported to resolve the ^{1}H NMR spectra of (*R*)-linalool and (*S*)-linalool.[134]

We first studied the linalool model system by analyzing a racemic standard (Fluka) by ^{1}H NMR and slowly adding Eu(hfc)$_3$ (Sigma) to the samples (data not shown). After addition, the two enantiomers began to exhibit differing ^{1}H NMR spectra, with the olefinic peaks shifting downfield. Further addition of the shift reagent also split the doublet into a doublet of doublets, showing the mixture was in fact racemic. Addition of a small amount of an (*R*)-linalool standard (Fluka) slightly increased the peak height of the doublet corresponding to that enantiomer. A racemic nerolidol standard (Sigma) was also tested using the same system, with similar results – the two enantiomers were evident, and the olefinic peaks began to split into a doublet of doublets with equal peak heights, signifying a racemic mixture.

The same procedure was then undertaken for the nerolidol purified from PMY1[pRS316Gal]. Addition of Eu(hfc)$_3$ shifted the allylic peak (5.232 ppm) further downfield, and eventually the doublet split into a doublet of doublets (J = 17 Hz and 3.6 Hz, Figure 4.3), suggesting that both enantiomers were present. Furthermore, the peak areas were equal, leading to the conclusion that nerolidol was present as a racemic mixture. Because most sesquiterpene synthases that catalyze formation of nerolidol produce a single enantiomer,[135-137] this result supported the possibility of nonenzymatic FPP hydrolysis.

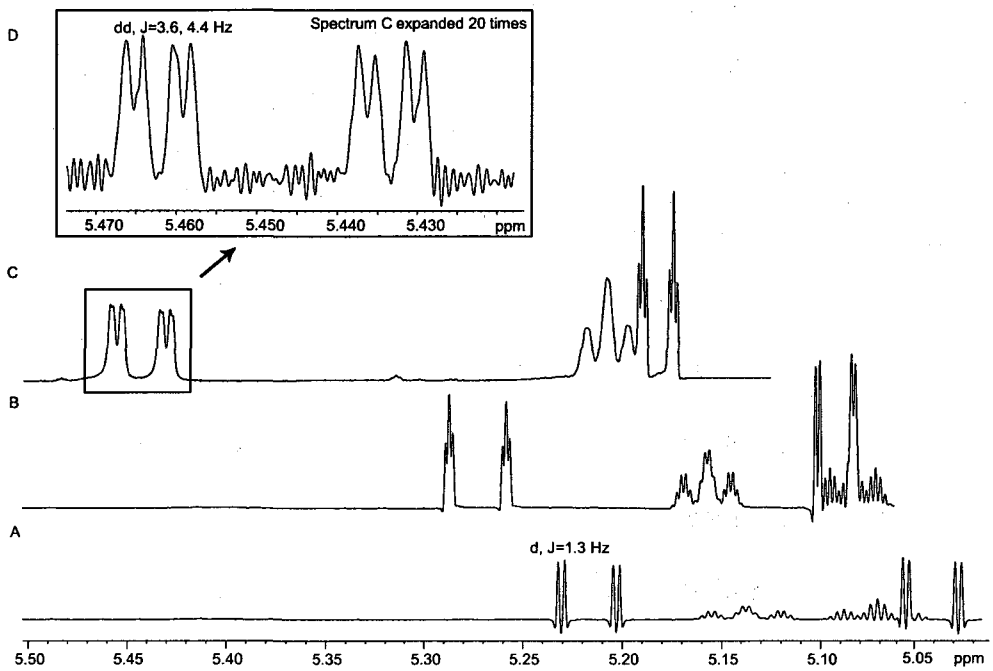

Figure 4.3. ^{1}H NMR spectra of nerolidol purified from PMY1[pRS316Gal] before and after addition of $Eu(hfc)_3$. Spectrum A is nerolidol prior to addition of the shift reagent and spectrum B shows the 5.232 doublet shifting downfield after addition of $Eu(hfc)_3$ and the doublet starts splitting into a triplet. Spectrum C shows a further downfield shift and the peaks splitting into a doublet of doublets. Spectrum D shows spectrum C after resolution enhancement, with fine splitting into a doublet of doublets (J = 17 Hz, 3.6 Hz).

As farnesol and nerolidol were isolated from the culture media, the media components were assessed for their possible influence on the sesquiterpene product profile in PMY1 with and without empty vector. Rich (YP) and synthetic (SC-U) media were examined, and yeast grown in SC-U media gave dramatically different results than previously seen (Figure 4.4). The synthetic media produced nerolidol as the major product and farnesol was less than half the level of nerolidol. Two formal possibilities could account for the difference in results between media: either some component present in YP promotes farnesol production, or some component in synthetic medium facilitates

nerolidol formation. To distinguish between these possibilities, a 1:1 mixture of YP and SC-U media was tested. The product profile resembled that of pure YP, with farnesol produced at levels seven times higher than nerolidol. The individual components of YP were also tested, to determine if one component could be directing farnesol production. Cultures grown in SC-U supplemented with either peptone or yeast extract also produced farnesol as the major product (Table 4.2).

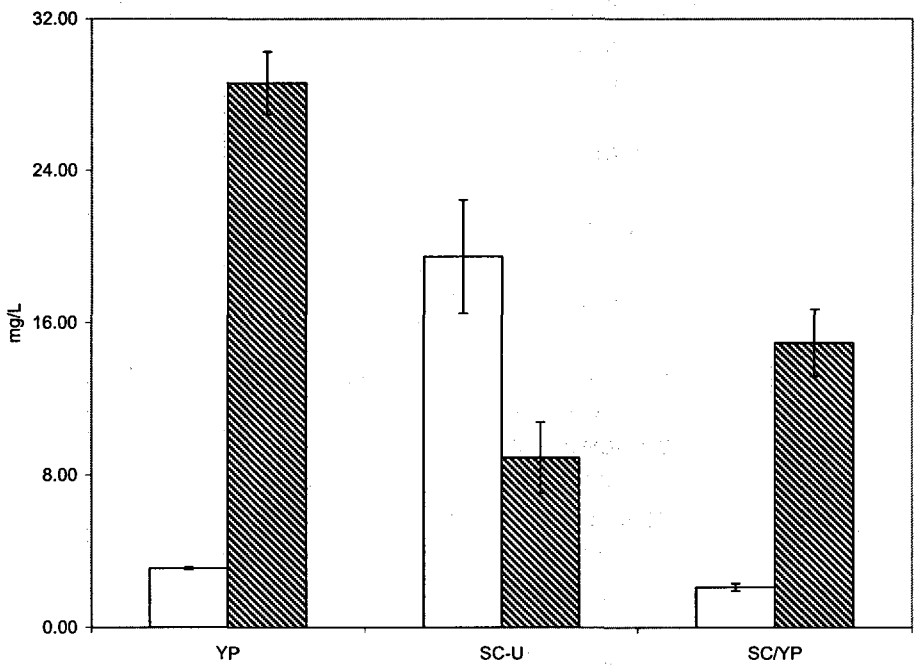

Figure 4.4. GC-FID quantitation results from PMY1 and PMY1[pRS316Gal] grown in YP, SC-U, or a 1:1 mixture of YP and SC-U. Farnesol is denoted by patterned columns while nerolidol is represented by white columns.

The possibility that pH changes could impact hydrolysis was considered early in this investigation, but both media had pH values too close to neutral (pH 5.5-7) to promote pyrophosphate elimination. An experiment was conducted to establish the extent to which yeast cultures become more acidic during growth. Cultures of PMY1 or PMY1[pRS316Gal] (50 mL) were grown in the three media conditions and the pH was

measured at initial inoculation, log-phase, and saturation phase (Table 4.2). Controls with no yeast inoculation were incubated in parallel.

Table 4.2. Cultures became more acidic 30 h and 48 h after inoculation.

Strain	Media	Initial pH	After 30 h	After 48 h	Major product
PMY1	YP	6.5-7	6	5.5	Farnesol
PMY1[pRS316Gal]	SC-U	5.5	5-5.5	4-4.5	Nerolidol
PMY1[pRS316Gal]	SC-U/YP	6	5.5-6	5	Farnesol
PMY1[pRS316Gal]	SC-U + yeast extract	6	5.8-6.1	Not measured	Farnesol
PMY1[pRS316Gal]	SC-U + peptone	6-6.5	6-6.5	Not measured	Farnesol
None	YP	7	6.5-7	6.5-7	N/A
None	SC-U	5.5	5.5	5.5	N/A
None	SC-U/YP	6	6-6.5	6-6.5	N/A
None	SC-U + yeast extract	6-6.5	6-6.5	Not measured	N/A
None	SC-U + peptone	6-6.5	6.5	Not measured	N/A

The SC-U culture started at a more acidic pH than the YP medium or mixtures of YP and SC-U, but all three media became more acidic as the yeast grew. The more neutral pH of the YP media appeared to direct FPP hydrolysis toward production of farnesol, while the more acidic SC-U increased formation of nerolidol (Table 4.3). Nerolidol formation becomes dominant when the final pH is below 5, so cultures that start with pH 6.5-7 have sufficient buffering capacity to remain above this level, accumlating farnesol as the major product. Most of the product accumulates once the medium has acidified, and so the final pH is probably more influential in directing the direction of FPP hydrolysis.

To preclude the possibility that an unidentified individual component of the media caused the difference in FPP hydrolysis product ratios, YP media was acidified to mimic

the pH of SC-U, and NaOH was added to SC-U media to mirror YP pH. The basified SC-U media led to results similar to that of YP; farnesol was isolated as the major product (Table 4.3). Acidic YP began resembling SC-U results as well, with a 1:1 mixture of farnesol and nerolidol produced when the pH was lowered to 5-5.5. These results confirmed that media pH was the controlling factor.

Table 4.3. pH-altered cultures produced different amounts of farnesol and nerolidol than the original cultures. All cultures were grown with galactose, ergosterol, and hemin. pH was measured at inoculation and after 48 h incubation.

Media	Initial pH	After 48 h	Farnesol/ Nerolidol
YP + MOPS	6-6.5	6-6.5	16.9
YP + acetic acid	5-5.5	5.0	1.22
SC-U + NaOH	7-7.5	4.5	18.4

It was previously thought that phosphatases were partially responsible for hydrolysis, and the PMY2 results reported here show that DPP1 has modest activity. A truncated alkaline phosphatase was isolated from *S. cerevisiae* that hydrolyzed FPP to farnesol,[77] and Poulter and Rilling[138] reported prenyltransferases capable of hydrolyzing geranyl diphosphate to the corresponding alcohols and hydrocarbons. It remains possible that another enzyme promotes FPP hydrolysis, but this work demonstrates that a majority of FPP hydrolysis is non-enzymatic.

These experiments establish that the pH of the medium influences whether FPP is hydrolyzed by S_N2 or S_N1/ S_N2' reactions. Acid in the medium probably does not directly catalyze the hydrolysis of intracellular FPP. Proton pumps prohibit the free diffusion of acid into the cell and intracellular pH is consequently rather independent of media pH. It

is conceivable that phosphatases are regulated by extracellular acid and therefore the enzymatic product profile depends on the acidity of the culture. However, this model would mandate that the phosphatase active site that generates nerolidol is electronically and sterically symmetrical, and it is difficult to conceive of a sufficiently strong selective pressure to promote racemic biosynthesis in this system. That the nerolidol is racemic is most readily explained by it being a nonenzymatic product arising from FPP being transported to the medium and undergoing an extracellular acid-mediated S_N1 reaction at the tertiary allylic center.

To test for the presence of minor sesquiterpene compounds and to provide further insights into the hydrolysis mechanisms, PMY1[pRS316Gal] products were purified by column chromatography with pentane and a mixture of pentane/ether (5:1). These fractions were analyzed by GC-MS and NMR, and several minor products were discovered (Figure 4.5).

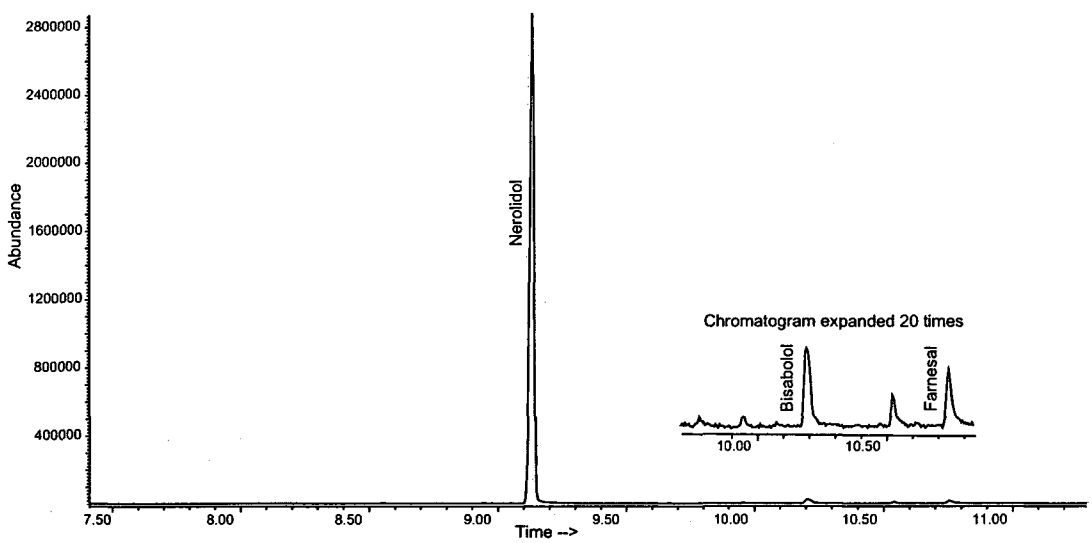

Figure 4.5. Total ion chromatogram of the PMY1[pRS316Gal] fraction containing nerolidol, α-bisabolol, and farnesal. The 10.1 to 10.8 minute range was expanded 20 times to show the minor compounds. PMY1[pRS316Gal] was grown in SC-U.

GC-MS analysis revealed a sesquiterpene alcohol peak at 10.2 min that had a retention time and fragmentation patterns that matched an authentic standard of α-bisabolol (See Appendix D for mass spectral data). ^{1}H NMR analysis confirmed the structure and indicated a pair of diastereomers. Limited amounts of material precluded chiral NMR studies to establish whether α-bisabolol was racemic or chiral. Previously, Carrau et al. reported that *S. cerevisiae* generates the monoterpene cyclic alcohol α-terpineol in under similar conditions,[139] suggesting non-enzymatic hydrolysis of geranyl pyrophosphate (GPP). α-Bisabolol most likely is derived from a similar mechanism using FPP. Hydrolysis of FPP or nerolidyl pyrophosphate (NPP) had been implicated previously in production of α-bisabolol in the cotton δ-cadinene synthase, where it was a minor enzymatic product.[140]

In addition, the pentane fractions showed two olefinic elimination products, α- and β-farnesenes, as established by comparison of the mass spectral patterns with those of authentic standards (See Appendix D for mass spectral data of farnesene). These compounds could form either through an E1 or an E2 elimination mechanism. These results demonstrate a complex product profile, with solvolysis leading to both substitution and elimination products. Additionally, farnesal and farnesyl acetate were also detected, based on their mass spectral patterns. These are readily rationalized as oxidized or acetylated metabolites of farnesol. Farnesyl acetate had a retention time and mass spectrum matching that of an authentic standard, while the farnesal mass spectrum was compared to that found in literature.[141] These minor components were found in very small levels, providing less than 2% of the product profile (Table 4.4). α-Bisabolol was

present at approximately 1%, whereas farnesal, farnesyl acetate, and farnesenes were found at levels less than 0.1%.

Table 4.4. Sesquiterpene products isolated from PMY1[pRS316Gal] grown in SC-U.

Compound	% Composition
Nerolidol	59.5%
Farnesol	39.5%
α-Bisabolol	1.05%
Farnesenes	0.08%
Farnesyl acetate	0.03%
Farnesal	0.02%

Figure 4.6 shows how the $C_{15}H_{26}O$ and $C_{15}H_{24}$ products are accessible upon ionization the farnesyl cation, which can be quenched at the primary or tertiary allylic positions to give farnesol and nerolidol, respectively. The α-bisabolol diastereomers arise from cation-olefin cyclization, followed by quenching through an S_N1-type mechanism. Products α-farnesene and β-farnesene could be generated from deprotonation through an E1 mechanism. Farnesal and farnesyl acetate are derived separately from farnesol.

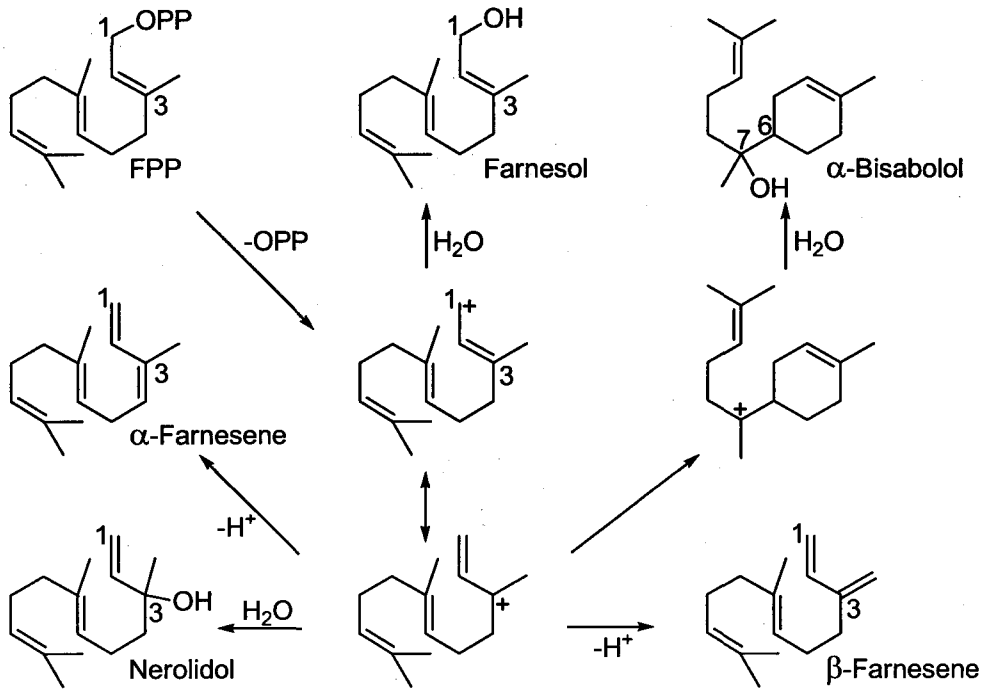

Figure 4.6. Proposed mechanism of PMY1[pRS316Gal] products from non-enzymatic FPP hydrolysis. S_N1 mechanisms lead to formation of farnesol, nerolidol, or bisabolol. α-Farnesene and β-farnesene could arise from E1 mechanisms.

However, Figure 4.6 is a simplification. It is difficult to envision α-bisabolol being formed by a process other than initial ionization, but alternative reasonable origins for other structures probably contribute to the product profile. The level of the primary alcohol farnesol is much too high for a pure S_N1 mechanism. Much of the farnesol probably arises from S_N2 attack, and an S_N2'-type reaction could form nerolidol, with water attacking at C-3. Some of the farnesol could arise from hydrolysis of the phosphate ester by nucleophilic attack on the α-phosphate. A bimolecular elimination could also occur to give α- and β-farnesenes through an E2 mechanism.

All of these processes can be envisioned as non-enzymatic reactions in the media, but only the argument for nerolidol is strongest because it was demonstrated to be

racemic. Although whether α-bisabolol is chiral could not be determined, its diastereomeric ratio and poor yields are consistent with non-enzymatic cyclization. Farnesol and the farnesenes are achiral and their structures consequently do not illuminate the fractions that are enzymatic or nonenzymatic in origin.

The extent to which enzymatic origins could contribute to the product profile can be considered in light of known enzymes and *S. cerevisiae* genomic sequence. The genome encodes other phosphatases which could generate farnesol. However, yeast lacks obvious homologs of known enzymes that generate farnesenes or α-bisabolol. Farnesenes can be rationalized as potential minor products of prenyl transferases or squalene synthase, which are known yeast enzymes that generate the farnesyl cation. Farnesenes seem likely by-products of processes that use the farnesyl cation, however the low levels produced here suggest that farnesenes may have evaded detection in previous work. That α-bisabolol could be a minor product of prenyltransferases seems less likely, because cyclization requires a coiled conformation that seems inaccessible in the extended active sites of these enzymes.

Metabolically engineered yeast are now used for commercial farnesol production, and the experiments described above establish that media manipulation can generate nerolidol. Farnesol and nerolidol are valuable compounds for the cosmetic and fragrance industries, and these alcohols have been used to enhance transdermal delivery of several drugs.[142,143] Additionally, farnesol and nerolidol have shown antimicrobial and limited antitumor activity.[144-148] These strains can be utilized to produce large amounts of these linear alcohols.

The discovery of substantial levels of non-enzymatic hydrolysis has substantial implications for sesquiterpene metabolic engineering. Metabolic engineering efforts to synthesize sesquiterpenes have focused on increasing accumulation of the FPP substrate. Manipulation of the yeast sterol biosynthetic pathway has produced sufficient FPP to aid in characterization of numerous sesquiterpene synthases, the enzymes that catalyze FPP cyclization, and P_{450} oxidases which further metabolize the hydrocarbon products.[88,131,149] Deletion of squalene synthase dramatically increased the amounts of sesquiterpene products isolated, but FPP hydrolysis still represents a large sink for FPP. Takahashi et al. previously established that expressing a terpene synthase in a DPP1 deletion does not substantially improvement in cyclization yields.[84] As shown above, neither DPP1 nor LPP1 deletion fundamentally alters FPP hydrolysis.

In metabolically engineered processes intended to biosynthesize cyclic terpenoids, FPP hydrolysis is a major competing outcome to cyclization. The experiments described herein point to severe limitations in abolishing FPP hydrolysis, suggesting that improving the efficiency of sesquiterpene synthases by improved expression technology or catalytic properties might be the next step needed to improve sesquiterpene cyclization yields.

The work presented here demonstrates a facile approach to altering media pH to preferentially generate farnesol or nerolidol as the major product. Deletion of DPP1 could aid in efforts to decrease the competing reaction of FPP hydrolysis, however, farnesol and nerolidol production cannot be completely abolished through phosphatase deletion alone.

Experimental Procedures

Construction of yeast strains. The yeast strains PMY1 and PMY2 were constructed by fellow graduate student Pietro Morlacchi. The BEJY14 strain was constructed by former lab member Dr. Beth Jackson.[150] PMY1 was transformed with the yeast expression vector pRS316Gal to yield a control strain PMY1[pRS316Gal] that grew on synthetic complete medium lacking uracil (SC-U).

Growth and isolation of sesquiterpene products from culture media. Single yeast colonies were used to inoculate 10-mL dextrose cultures of either YP (PMY1, PMY2, or BEJY14) or SC-U (PMY1[pRS316Gal]). Because these strains did not have intact sterol biosynthetic pathways, cultures were supplemented with heme and ergosterol. Galactose cultures (100 mL) were inoculated with the saturated 10-mL dextrose cultures. Once saturated, these cultures were transferred to 250-mL flasks containing methanol-activated hydrophobic resin (Diaion-20, Supelco) and were incubated at 30 °C with shaking overnight. Cultures were filtered through Kontes columns and the cells and media were set aside. The resin remaining in the column was washed with 3 volumes of ethanol and the internal standard longifolene (100 μg) was added to the extracts. The ethanol was then partitioned between one-half volume of water and three one-half volumes of pentane. The combined pentane extracts were concentrated by rotary evaporation and transferred to GC vials with 1 mL pentane. Ten percent of each extract was used for GC-MS and GC-FID analysis. The peak areas of farnesol and nerolidol were compared to that of longifolene to determine the amount of products

isolated from each culture. Triplicate cultures ensured more accurate numbers were obtained.

Isolation of sesquiterpenes from cell pellets. The 100-mL galactose cultures used for resin extractions were centrifuged (3000 rpm, 10 min) after filtering through the Kontes columns. The resulting cell pellets were subjected to saponification in 5 mL 10% KOH in 80% EtOH with 200 mg BHT (butylated hydroxytoluene, an antioxidant). The solutions were heated in a 70 °C water bath for 2 h and allowed to cool to room temperature. Distilled H_2O (5 mL) was added and the solutions were extracted with 3 × 15 mL pentane, and the solvent was then removed in vacuo. The residues were transferred to GC vials with 1 mL pentane and 100 μg longifolene was added prior to analysis by GC-MS and GC-FID.

Isolation of minor products in PMY1[pRS316Gal]. A 1-L galactose culture of PMY1[pRS316Gal] was grown in SC-U and the organic compounds were isolated as described above. The resin extract was chromatographed over a silica gel column using a 5:1 pentane/ether mixture. Finally 100% ether was used to elute any remaining compounds. Each fraction was analyzed by TLC and GC-MS, and possible minor products were tentatively identified.

Determination of nerolidol stereochemistry using a chiral shift reagent. A lanthanide chiral shift reagent was utilized to determine its effect on ^1H NMR chemical shifts of the nerolidol product, following the protocol outlined by Wilson et al.[134] The

alcohol group in linalool has the same steric and electronic environment as nerolidol, and linalool standards were first tested to determine how the reagent affects the chemical shifts.

Europium tris[2-(heptafluropropylhydroxymethylene)-(+)-camphorate] ($Eu(hfc)_3$) was diluted to a 4.7 nM solution in $CDCl_3$ and was slowly added to the racemic linalool standard. As more $Eu(hfc)_3$ was added, the peaks began shifting downfield and split the enantiomers. By adding (-)-linalool standard (Fluka), we were able to overlay signals for the shifted enantiomer.

A racemic nerolidol standard (Sigma) was also tested in this system to ensure that it had the same behavior as the linalool standard, and the olefinic peaks did indeed shift further downfield and split slightly when shift reagent was added. The nerolidol isolated and purified from the yeast was diluted in $CDCl_3$ and analyzed by ^1H NMR prior to addition of the shift reagent. $Eu(hfc)_3$ was slowly added and the olefinic peaks were shifted from 5.232 ppm to 5.462 ppm.

Culture media effects on sesquiterpene production. The yeast strains PMY1 and PMY1[pRS316Gal] were grown in galactose with different nitrogen sources to determine whether or not media composition played a role in FPP hydrolysis. PMY1 was grown in rich medium (YP) and PMY1[pRS316Gal] was grown in synthetic medium (SC-U), a 1:1 mixture of YP and SC-U, or SC-U supplemented with either yeast extract or peptone. Once saturated, the cultures were transferred to activated resin and extractions were carried out as described above. The media extracts were quantitated by GC-FID.

Media pH effects on sesquiterpene production. Looking at the pH at various points between inoculation and saturation of the cultures and the corresponding controls provided a more comprehensive idea of how pH changed with incubation time. Galactose cultures (50 mL) were inoculated from dextrose cultures, and 50-mL galactose control cultures containing no yeast were also prepared. pH measurements were recorded at time of inoculation and after 30 and 48 hours incubation. Cultures were then harvested and resin was extracted as described above.

An experiment to decrease the acidity of YP to that of SC-U was then carried out. A 100-mL galactose culture of YP was prepared, and NaH_2PO_4 was added for a final concentration of 180 mM to decrease the pH of the media to approximately 5.5. Another 100-mL YPGHE culture was prepared, and MOPS (buffering range pH 6.5-7.9) was added to a final concentration of 50 mM, with pH measured at 6.5-7.0. Alternatively, the pH of a 100-mL YPGHE culture was adjusted to 5.0-5.5 with the addition of glacial acetic acid. pH of these cultures was monitored over the incubation time.

Additionally the pH of SC-U was increased to mimic that of YP. First, a small amount of NaOH was added to 50 mL SCG-UHE prior to inoculation with PMY1[pRS316Gal]. The pH of the culture increased from 5.5 (prior to NaOH addition) to 7.0-7.5 (after addition). Another 50-mL culture of PMY1[pRS316Gal] was grown in SCG-UHE after addition of a small pellet of NaOH and 1 g MOPS buffer, with the pH increasing from 5.5 to 6.0-6.5. Finally, a 50-mL culture was supplemented with 5 M NaOH to pH 7.0-7.5. pH of all cultures was monitored throughout incubation and at the time of harvesting.

Characterization of (±)-(6*S*,7*S*)-α-bisabolol and (±)-(6*S*,7*R*)α-bisabolol. ^{1}H NMR spectra were measured on a Varian 800 MHz spectrometer. The α-bisabolol ^{1}H NMR chemical shifts matched those reported in literature.[151]

Table 4.5. (±)-(6*S*,7*S*)-α-Bisabolol ^{1}H NMR chemical shifts, comparing literature and observed values. An asterisk indicates the peak was partially covered by a nerolidol peak so accurate values could not be determined.

Atom #	Literature value	Observed value
H-2	5.37 (br s)	5.375, m
H-10	5.12 (t, J = 7 Hz)	5.131, t, J = 7.1, 1.4 Hz
H-12	1.68	1.68*
C-3 Me	1.64	1.652
C-7 Me	1.10	1.110
C-11	1.62	1.620

Table 4.6. (±)-(6*S*,7*R*)-α-Bisabolol ^{1}H NMR chemical shifts, comparing literature and observed values. An asterisk indicates the peak was partially covered by a nerolidol peak so accurate values could not be determined.

Atom #	Literature value	Observed value
H-2	5.39 (br s)	5.39, m
H-10	5.12 (t, J = 7 Hz)	5.131, t, J = 7.1, 1.4 Hz
H-12	1.68	1.68*
C-3 Me	1.64	1.652
C-7 Me	1.13	1.139
C-11	1.62	1.620

Chapter 5

Accumulation of triterpene substrates oxidosqualene and dioxidosqualene

Introduction

Triterpenoids are derived from three different acyclic 30-carbon precursors – squalene, oxidosqualene (OS), or dioxidosqualene (DOS).[152] *S. cerevisiae* produces all three of these compounds, and squalene and OS are intermediates in the sterol biosynthetic pathway (Figure 5.1). Squalene is oxidized by squalene epoxidase (ERG1) to form (3*S*)-oxidosqualene.[43] When yeast with a native sterol biosynthetic pathway experience normal vegetative growth, oxidosqualene does not accumulate to any appreciable extent but is cyclized by lanosterol synthase (ERG7) to lanosterol, the first committed intermediate in sterol biosynthesis.[52] When lanosterol synthase activity becomes limiting, OS can accumulate and then re-enter the ERG1 active site for further epoxidation. The resultant molecule (3*S*,22*S*)-dioxidosqualene[50] is not utilized in the formation of ergosterol but is an intermediate in the production of epoxysterols, which play a regulatory role in sterol biosynthesis.[55]

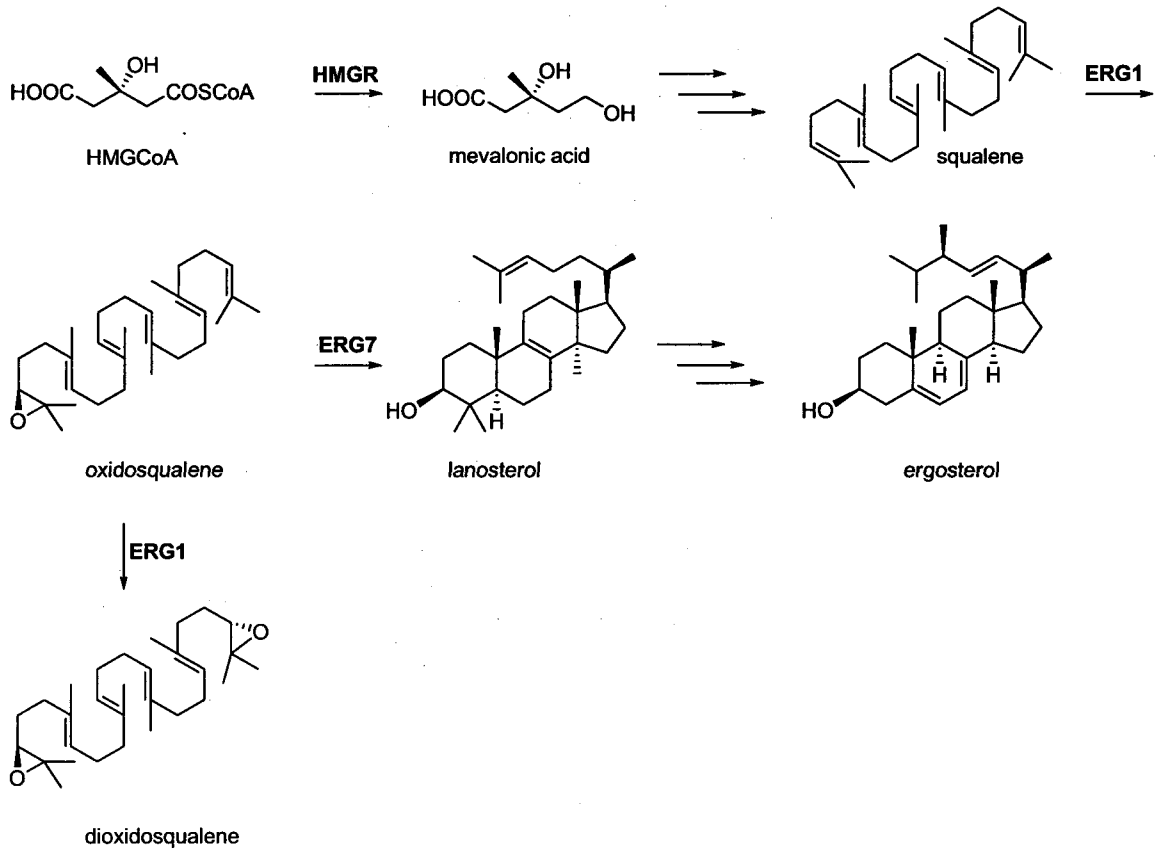

Figure 5.1. Yeast sterol biosynthetic pathway and formation of DOS. ERG1 epoxidizes squalene to form OS, which can then re-enter the active site to produce DOS.

Results and Discussion

Lanosterol synthase mutants such as SMY8[120] have been the preferred expression hosts for in vivo and in vitro systems to biosynthesize triterpenes.[98,99,101,102,153] The absence of lanosterol synthase abolishes native cyclization to promote the accumulation of OS. Two SMY8 derivatives were constructed in an attempt to increase accumulation of the triterpene precursors OS and DOS. The yeast strain ABY1 was generated by overexpressing a solubilized form of HMG-CoA reductase (trHMG1)[82] in SMY8 to increase carbon flux through the biosynthetic pathway (Table 5.1). To promote DOS

rather than OS biosynthesis, we constructed a strain (ABY2) that overexpressed the yeast squalene epoxidase ERG1.

Table 5.1. Summary of constructed yeast strains. All strains are MAT**a** and *erg7*::HIS3 *hem1*::TRP1 *ura3-52 trp1-Δ63 his3-Δ200 ade2* GAL^+.

Strain	Characteristics
SMY8[pRS305Gal]	Control strain
ABY1	*trHMG1::LEU2*
ABY2	*ERG1::LEU2*

The ability of these strains to accumulate acyclic triterpenoids was investigated using galactose as the carbon source. The SMY8 parent and the two derivative strains biosynthesize the triterpene precursors squalene, OS, and DOS without need for inducing medium. The genetic modifications that yielded ABY1 and ABY2 are driven by galactose-inducible promoters and thus required galactose to induce expression. The SMY8[pRS305Gal] control was also induced with galactose to ensure consistency between cultures. The amount of inoculum can be important. Excessively large inocula give poorer yields because of insufficient generations in inducing medium, whereas insufficient inocula can yield cultures that grow so long under inducing conditions that poorly-expressing mutant derivatives (which can have a selective advantage) can take over the culture. A 1% inoculum struck a good balance.

For each strain, cultures were harvested at log-phase and the cell pellets were saponified and extracted. The resulting non-saponifiable lipids (NSL) were TMS-derivatized overnight and quantitated by GC-FID, with the peak areas of squalene, OS, and DOS compared to that of TMS-epicoprostanol as internal standard (Table 5.2). Several observations were made from the cell pellet data. SMY8[pRS305Gal] served as

the control strain, producing 0.5 mg/L total triterpenes, with four times more DOS than OS produced. On the other hand, ABY1 accumulated twice as much OS as DOS with 10 mg/L total linear triterpenes. This strain was dramatically different than the other two strains, as it was the only strain producing more OS than DOS.

Table 5.2. Quantitation of NSL based on GC-FID integration. Product areas were compared to that of TMS-epicoprostanol. Each yield was based on triplicate cultures. N.D. indicates the product was <1% of the major product.

Strain	Squalene (mg/L)	OS (mg/L)	DOS (mg/L)	DOS/ OS
SMY8[pRS305Gal]	N.D.	0.11 ± 0.03	0.41 ± 0.10	3.70
ABY1	0.35 ± 0.05	6.78 ± 0.56	3.39 ± 0.35	0.50
ABY2	N.D.	0.04 ± 0.01	0.19 ± 0.01	5.04

trHMG1 overexpression sharply increased the production of acyclic triterpenes; the ABY1 strain accumulated approximately 20 times more squalene and derivatives than did SMY8. This strong effect confirms that mevalonate production limits OS biosynthesis in SMY8. The decreased DOS/OS ratio in ABY1 is readily rationalized in light of the introduced changes. The overexpression of trHMG1 introduces more carbon into the sterol biosynthetic pathway and thereby promotes squalene biosynthesis. The influx of new squalene apparently exceeds the native ability of squalene epoxidase to consume it, and a substantial portion of the total material resides as the monoexpoxide.

ERG1 overexpression had modest influence on OS and DOS production during log-phase. The ABY2 strain had a slightly higher DOS/OS ratio (5:1), compared to SMY8[pRS305Gal] (4:1).

To investigate whether triterpenes were secreted into the culture medium, the log-phase culture media were extracted and the acyclic products were quantitated by GC-FID

(Table 5.3). The data showed that the squalene derivatives accumulated in the media at levels up to five times higher than that isolated from the cell pellets. From this data, the more polar DOS appears to be preferentially secreted into the medium instead of OS. Because of this, the results described for the cell pellets were not a comprehensive enough picture for the roles of trHMG1 and ERG1 overexpression in SMY8.

Table 5.3. Quantitation of media extracts based on GC-FID integration. Product areas were compared to that of TMS-epicoprostanol. Each yield was based on triplicate cultures. N.D. indicates squalene was not detected within a 1% detection limit.

Strain	Squalene (mg/L)	OS (mg/L)	DOS (mg/L)	DOS/ OS
SMY8[pRS305Gal]	N.D.	0.16±0.03	1.05±0.24	6.59
ABY1	N.D.	0.12±0.02	1.25±0.11	10.6
ABY2	N.D.	0.03±0.01	1.65±0.28	61.7

The media from SMY8[pRS305Gal] and ABY1 cultures contained similar amounts of OS and DOS, with about 1.3 mg/L total triterpenes isolated from both strains. Therefore the cell pellet data is more definitive in understanding the role of trHMG1 overexpression on squalene derivative biosynthesis. The ABY2 strain overexpressing ERG1 accumulates more DOS and less OS in the medium than the parent strains does, with a DOS/OS ratio 10 times higher than that of SMY8. From these results the media illuminated the effect of ERG1 overexpression more strongly than the cell pellet data.

In order to better understand the trends between the three strains, the total amounts of triterpenes produced by these strains were calculated by combining the cell pellet and media data (Table 5.4), which show substantial differences between the engineered strains and the parent SMY8 strain. ABY1 accumulated the most triterpene

precursors, with nearly 12 mg/L total triterpenes, whereas SMY8[pRS305Gal] and ABY2 accumulated similar amounts of OS and DOS.

Table 5.4. Total triterpenes from each strain, isolated from both cell pellets and media. Quantitation is based on GC-FID integration and numbers are reported in mg/L.

Strain	Squalene (mg/L)	OS (mg/L)	DOS (mg/L)	DOS/ OS
SMY8[pRS305Gal]	0.04±0.02	0.28±0.05	1.59±0.28	5.65
ABY1	0.35±0.05	6.90±0.56	4.68±0.27	0.68
ABY2	N.D.	0.06±0.01	1.83±0.29	28.2

The three strains exhibited significant differences in DOS/OS ratios. ABY1 produced almost twice as much OS than DOS. The other two strains produced more DOS than OS, with SMY8 having the smaller ratio. ABY2 has the largest DOS/OS ratio of the three strains, as intended; its overexpressed ERG1 facilitates the oxidation of OS to DOS.

In addition to the log-phase measurements, cultures were also harvested at saturation and the acyclic precursors were quantitated. Media and cell pellets from SMY8[pRS305Gal], ABY1, and ABY2 were studied, and the results were quite different from those observed in the log-phase cultures. The cell pellet data was consistent among all strains, with ratios of DOS/OS comparable between log-phase and saturation cultures. The media extracts again contained a majority of the acyclic compounds and demonstrated an increase in DOS production over incubation time.

Lengthening culture times increased the DOS/OS ratios because precursor biosynthesis stops when carbon is exhausted, but oxidation continues. The amount of acyclic precursor per liter and the DOS/OS ratios remain similar throughout log-phase SMY8[pRS305Gal] growth. The final culture was harvested after saturation, resulting in

the largest cell pellet and the largest DOS/OS ratio. The greater mass of acyclic triterpenes at saturation indicates that the 30-carbon molecules were converted from earlier precursors (e.g., FPP or squalene) at saturation (Figure 5.2). Over time more DOS was produced in the strain, increasing the DOS/OS ratio and indicating that further oxidation took place after saturation.

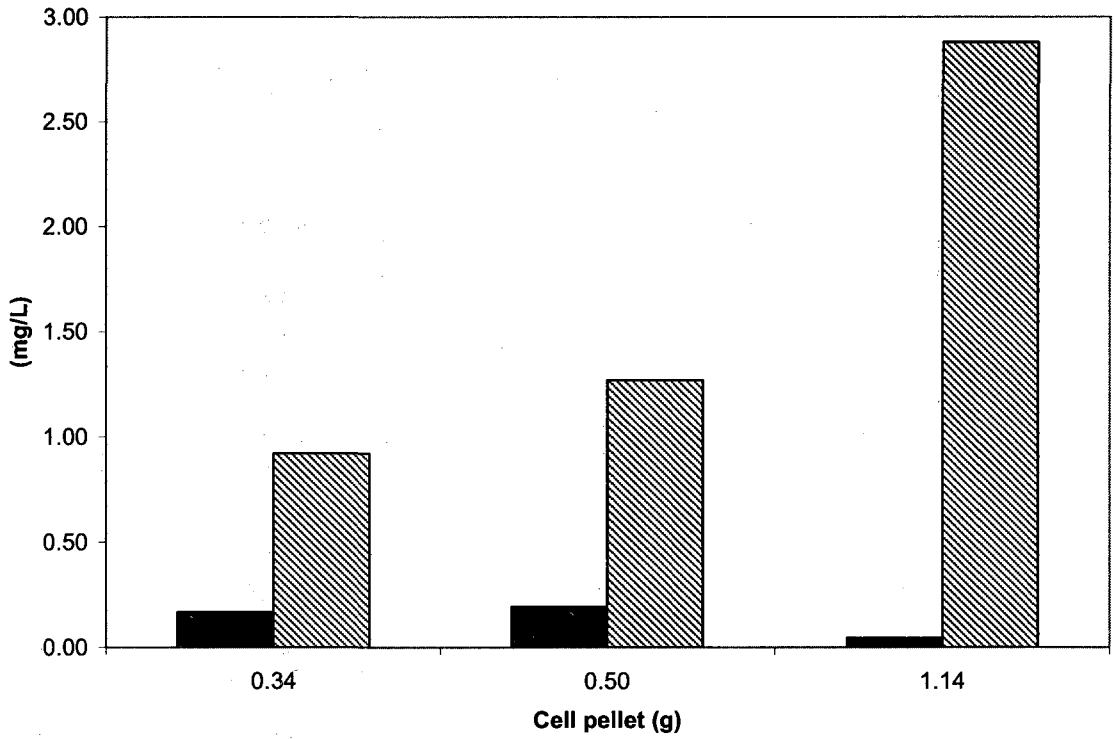

Figure 5.2. GC-FID quantitation of SMY8[pRS305Gal] media extracts harvested at varied growth stages. Cultures ranged from one day growth (0.34 g) to log-phase (0.50 g) to saturation (1.14 g). The black columns represent OS production in mg/L, while the patterned columns represent DOS production in mg/L.

Similar results were observed with ABY2 (Figure 5.3), as increasing amounts of DOS accumulated as cultures were incubated longer. The amount of OS also decreased with incubation time, further increasing the DOS/OS ratio. Because the cultures were inoculated from dextrose-containing cultures, galactose induction did not occur

immediately.[154] Once the dextrose was consumed, galactose was utilized by the yeast and the ERG1 gene was finally turned on to produce more OS and DOS. Harvesting cultures too early, as shown in the first culture, meant less material was isolated from the early culture.

Comparing the results from SMY8[pRS305Gal] and ABY2, the SMY8[pRS305Gal] cultures were allowed to incubate longer and thus had a larger amount of total triterpenes, in comparison to the ABY2 cultures. ABY2 accumulated almost 2 mg/L OS and DOS, while SMY8[pRS305Gal] produced nearly 3 mg/L, likely explained in the longer incubation time.

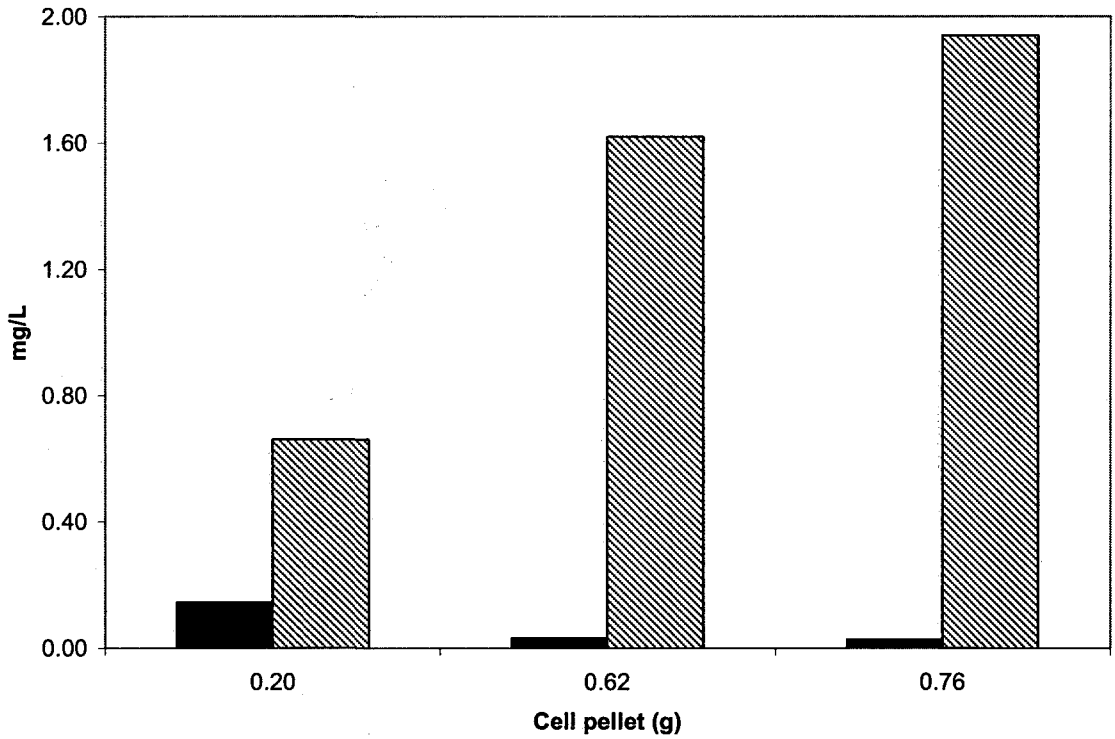

Figure 5.3. ABY2 media extract quantitation data from cultures harvested at different growth stages. The first culture was harvested after only one day, while the 0.62 g cell pellet was from a 2-day growth period (log-phase). The 0.76 g cell pellet was obtained after 3 days growth (saturation). The black columns represent OS production (mg/L) while the patterned columns represent DOS accumulation (mg/L).

Finally, the medium from the ABY1 strain was studied, and again the amount of DOS increased over a longer incubation time (Figure 5.4). Unlike the other strains, ABY1 had an increase in both OS and DOS over time. Additionally, at log-phase the amount of triterpenes accumulating was nine times lower than that observed for the saturated culture. The final culture harvested accumulated nearly 10 mg/L total triterpenes, the most isolated from any of these strains at any growth stage.

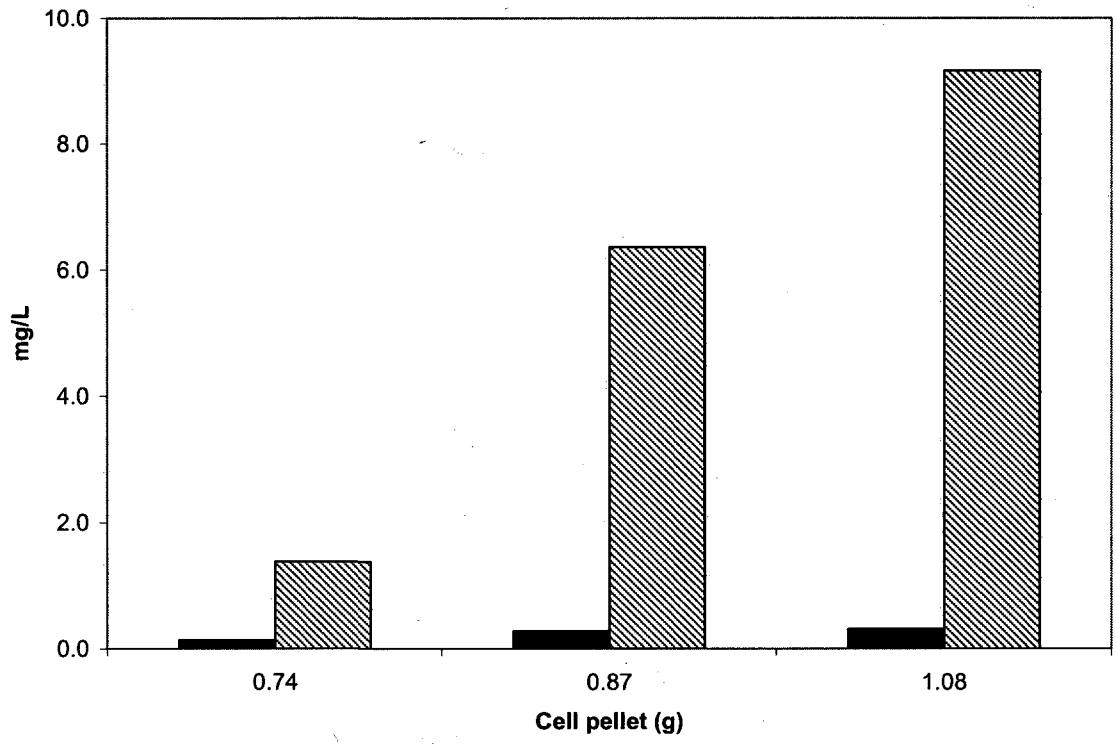

Figure 5.4. GC-FID quantitation data for ABY1 cultures harvested after different incubation times. The first two (0.74 g and 0.87 g) were harvested during log-phase, while the 1.08 g cell pellet is from a saturated culture after 3 days growth. Black indicates OS accumulation (mg/L) while the patterned columns represent DOS isolated from the media (mg/L).

Careful consideration must be given to which growth stages of the yeast to examine to best understand the implications of manipulating these enzymes. It is relatively difficult to draw conclusions from OS/DOS ratios at saturation because OS can

re-enter the ERG1 active site and be epoxidized on its distal terminus to convert OS to DOS. More DOS accumulated the longer cultures were allowed to incubate, thus the longer incubation time skewed the DOS/OS ratios of the yeast strains, with more DOS isolated than that observed in the log-phase cultures. Log-phase cultures are consequently essential in establishing the relevance of ERG1 and trHMG1 to metabolic engineering conditions in which those the acyclic precursors would be cyclized in vivo. However, these log-phase cultures do not give a clear understanding of the ultimate amounts that could accumulate in a metabolically engineered system as would be seen if the material were harvested at saturation; the cell mass is less than a saturated culture, and the product yields are also low due to a lag-time in galactose induction. Quantitation of saturated cultures are therefore necessary for determining the total amounts of squalene derivatives the strains are capable of producing.

Previously, most triterpene studies focused on triterpene products accumulating within the cells. This work demonstrates the necessity of studying the culture media, as up to 10 times more products were isolated from the media compared to the cell pellets.

The full potential of the ABY2 strain was studied by growing triplicate 1-L cultures in optimized conditions described previously[155] and harvesting at saturation. Because the ERG1 overexpression cassette was integrated, plasmid selection was not necessary and cultures could be grown in rich medium (YP with 2% adenine and 4% galactose). These nitrogen source optimizations coupled with higher galactose concentration provided greater cell mass and consequently more OS and DOS were produced from each culture. To quantitate the DOS isolated from the resin, the peak areas of DOS and TMS-epicoprostanol from GC-FID analysis were compared. The

optimization increased DOS yield 10-fold compared to the 100-mL cultures, yielding 17.0 ± 0.9 mg/L DOS and 0.4 ± 0.1 mg/L of OS. This gave a DOS/OS ratio of 46.2.

To obtain the most accurate quantitation results, specific structural properties of the acyclic triterpenes needed to be considered. OS and DOS are difficult to discriminate between (and consequently, to quantitate) by ^{1}H NMR because both have epoxide termini with essentially indistinguishable ^{1}H NMR signals. Similarly, squalene and OS share olefinic termini with very similar ^{1}H NMR signals. GC-based analyses of these compounds could be complicated by the limited thermal stability of the strained epoxides on the GC column, and that they may not have GC-FID and GC-MS responses identical to that of epicoprostanol. In order to determine what, if any, effect the structure of the compounds could have on the quantitation results, mixtures of standards were prepared (in 1:1 ratios in xylenes) and analyzed by GC-FID, GC-MS, and ^{1}H NMR. Six mixtures were prepared: squalene + OS, squalene + DOS, squalene + epicoprostanol, OS + DOS, OS + epicoprostanol, and DOS + epicoprostanol.

Because the yeast extracts were quantitated on the GC-FID by comparing peak areas of the products to the peak area of TMS-epicoprostanol (abbreviated as TMS-epi), the mixtures of standards were derivatized and analyzed by GC-FID to determine the relative responses of squalene, OS, DOS, and epicoprostanol (Table 5.5). Also essential was to determine how derivatization affects the FID response. The trends were verified by analyzing the same samples on GC-MS. TMS-derivatization converts the alcohol moiety in epicoprostanol to the corresponding TMS-ether (Figure 5.5), making the compound more volatile and preventing much of the tailing observed in non-derivatized triterpene alcohols. This tailing prevents accurate integration of the epicoprostanol peak,

skewing the results. TMS-derivatization adds a silicon atom and three methyl groups, and it was essential to establish how this change would affect FID response. In addition, it was not known whether the reagents used for TMS-derivatization would affect OS or DOS. Although these compounds do not contain the necessary alcohol groups, it is conceivable that TMS-derivatization conditions could lead to OS and DOS degradation, or that the epoxides would thermally decompose on the column. The mixtures not containing epicoprostanol were derivatized as well to study what, if any, effect could occur.

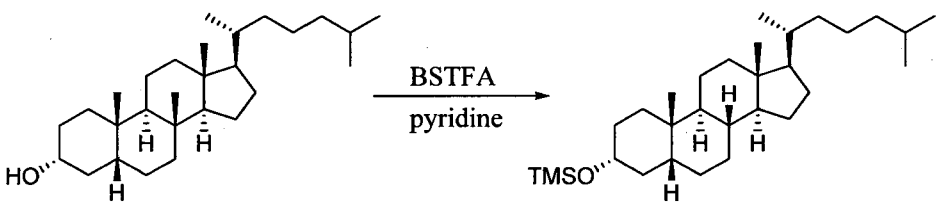

Figure 5.5. TMS-derivatization of epicoprostanol.

Table 5.5. GC-FID quantitation of triterpene standards before and after derivatization.

Mixture	Before TMS-derivatization	After TMS-derivatization	Percent difference	Effect on quantitation ratio
Squalene/TMS-epi	0.99	0.75	-24%	Enhances
OS/Squalene	1.16	1.11	+4.0%	Negligible
OS/TMS-epi	1.10	0.83	-25%	Enhances
DOS/TMS-epi	0.85	0.63	-26%	Enhances
DOS/Squalene	0.88	0.86	+2.3%	Negligible
DOS/OS	0.78	0.77	+1.3%	Negligible

The effect of derivatization was insignificant for the linear triterpenes, as the ratios of OS/squalene, DOS/squalene, and DOS/OS only increased 2-4% compared to the

non-derivatized ratios. The higher ratios observed for these three mixtures could be attributed to small amounts of degradation from the derivatization procedure.

The mixtures containing epicoprostanol, on the other hand, showed dramatically different ratios when comparing non-derivatized to derivatized samples, with all ratios decreasing 25% compared to the non-derivatized ratio. TMS-epicoprostanol produced a much sharper peak on GC-FID and GC-MS, with less peak tailing, which made peak integration much more accurate than non-derivatized epicoprostanol. Additionally, the TMS-epicoprostanol has more atoms per sample than non-derivatized epicoprostanol, dramatically changing the FID response. The TMS-derivatized samples were used for all quantitation experiments. Because of the large differences in the ratios, however, a correction factor was necessary to provide the most accurate numbers.

To verify these results, the same samples were analyzed by 600 MHz ^{1}H NMR. Comparisons were made primarily in the upfield methyl regions (see experimental procedures for chemical shift information). NMR integration did not provide definitive ratios for all mixtures due to contaminations present in some samples causing inaccurate integration. However, the squalene/DOS and DOS/epicoprostanol mixtures gave clear results that were consistent with the initial mass measurements and were used as a standard with which to develop conversion factors for the GC-FID results (Table 5.6). Based on these two analytical methods, we determined that the epoxide moiety in DOS is not appreciably decomposed by the derivatization reagents or by 30 min exposure to the 260 °C column. Squalene, OS, and DOS do not have substantial differences in thermal stability on GC-FID, and their FID responses are within experimental error of one another.

Table 5.6. Comparison of ratios determined from GC-FID and ^{1}H NMR.

Mixture	Before TMS-derivatization	After TMS-derivatization	NMR calculated ratio	Final correction factor
DOS/squalene	0.88	0.86	0.90	Negligible
DOS/epicoprostanol	0.854	0.628	0.83	1.32

The amounts of DOS were thus recalculated for each yeast strain, taking the correction factor into account. Minor experimental errors, such as weighing (~2% error) and measuring of solvents (~1%) were not included in the correction factors, as these issues were present for all samples prepared. GC-FID, GC-MS, and NMR analyses revealed impurities in squalene, OS, DOS, and epicoprostanol. OS had <1% impurities whereas the other three standards contained approximately 5% impurities. These were accounted for in a correction factor. First, the NSL values from the log-phase cultures were recalculated (Table 5.7). The media extracts were similarly recalculated (Table 5.8).

Table 5.7. Corrected quantitation data of log-phase NSL. Results are based on GC-FID integration, with product areas compared to that of TMS-epicoprostanol. Each yield was based on triplicate cultures and includes a correction factor (Linear triterpenes/TMS-epicoprostanol = 1.32).

Strain	Squalene (mg/L)	OS (mg/L)	DOS (mg/L)	DOS/ OS
SMY8[pRS305Gal]	N.D.	0.15 ± 0.04	0.54 ± 0.13	5.34
ABY1	0.46 ± 0.07	8.95 ± 0.74	4.48 ± 0.46	0.50
ABY2	N.D.	0.05 ± 0.01	0.25 ± 0.01	5.39

Table 5.8. Corrected quantitation values for media extracts. Quantitation was based on GC-FID integration. Product areas were compared to that of TMS-epicoprostanol. Each yield was based on triplicate cultures and includes a correction factor (Linear triterpenes/TMS-epicoprostanol = 1.32).

Strain	Squalene (mg/L)	OS (mg/L)	DOS (mg/L)	DOS/ OS
SMY8[pRS305Gal]	0.02 ± 0.01	0.21 ± 0.04	1.39 ± 0.315	8.90
ABY1	N.D.	0.16 ± 0.02	2.18 ± 0.187	14.3
ABY2	N.D.	0.04 ± 0.01	2.17 ± 0.374	66.0

It was also useful to study the effects of differing chemical functionalities present when measuring product accumulation, as demonstrated here. Squalene, OS, and DOS demonstrated similar behavior on GC-FID, whereas TMS-epicoprostanol had differences in integration caused by extra atoms and sharper peaks, and these were accounted for by including a correction factor. The same correction factor (1.32) was used to recalculate yields for both OS and DOS accumulation.

While the correction factor did not change the DOS/OS ratios significantly, the correction did enhance the overall triterpene substrate yields. In the ABY1 strain, the total amount of OS and DOS accumulating in both the cell pellets and media was nearly 17 mg/L, compared to the original calculation of almost 12 mg/L. ABY1 accumulated eight times more material than the parent SMY8 strain, which was recalculated to accumulate 2.3 mg/L of OS and DOS between the cell pellets and media. ABY2 still accumulated similar amounts of triterpene precursors compared to SMY8, but the DOS/OS ratio was slightly higher than that previously observed.

The work presented here shows the remarkable potential of metabolic engineering for triterpene biosynthesis. By overexpressing the rate-limiting enzyme trHMG1 in the SMY8 parent strain, the amounts of epoxidized triterpene substrates accumulating increased 10-fold, producing more OS and DOS than the parent strain. This is consistent

with previous data in which overexpression of trHMG1 increased sesquiterpene and diterpene production.[35,82] Overexpression of the yeast ERG1 did not increase the amount of substrate accumulating in comparison to SMY8, but the higher epoxidation capacity of ABY2 increased the DOS/OS ratio dramatically, with nearly 80 times more DOS observed than OS.

The strains presented here produce large amounts of the triterpene substrates OS and DOS and are useful tools for studying OS or DOS cyclization. For example, the OS produced by ABY1 is the substrate for oxidosqualene cyclases (OSCs). Expression of OSCs in ABY1 would therefore facilitate the characterization of novel enzymes. The higher substrate levels should increase triterpene product yields, providing more material for structural identification.

Additionally, OS and DOS can be accepted by many OSCs, including *Saccharomyces* lanosterol synthase,[53,54] *Arabidopsis* thalianol synthase,[153] and *Arabidopsis* lupeol synthase.[103] Utilizing strains that can produce larger amounts of both precursors can therefore aid in characterization of products derived from both OS and DOS cyclization. Several epoxy triterpenes have been definitively or putatively identified as DOS cyclization products,[153,156-159] and the wealth of products from this alternate substrate is still not fully realized. Epoxylanosterol and oxysterols play a role in sterol homeostasis in animals and fungi and have potential medicinal utility.[160-162] Additionally these compounds could be utilized by plants as potential defense compounds. Improved means to uncover useful structures and interesting activities will expedite progress in these areas.

Experimental Procedures

Construction of yeast strain ABY1. The plasmid pEH12.4, containing the N-terminal truncated HMG1 gene[82] in the yeast expression vector pRS305biGal, was made by former lab member Dr. Gia Fazio. This plasmid was linearized with *Bst*E II and integrated into the lanosterol synthase mutant SMY8 at the leu2 locus using the lithium acetate method.[119] Transformants were selected on synthetic complete medium lacking leucine, yielding the ABY1 yeast strain (MAT**a** *erg7*::*HIS3 hem1*::*TRP1 trHMG1*::*LEU2 ura3-52 trp1-Δ63 his3-Δ200 ade2* GAL$^+$).

Construction of ERG1 overexpression strain. The plasmid pGCF6.0 containing *ScERG1*, made by former member Dr. Gia Fazio, is in the integrative yeast expression vector pRS305Gal.[155] The *Bsr*G I-linearized plasmid was integrated into SMY8 at the *leu2* locus, as discussed above. Transformants were selected on synthetic complete medium lacking leucine. The resulting strain was named ABY2 (MAT**a** *erg7*::*HIS3 hem1*::*TRP1 ERG1*::*LEU2 ura3-52 trp1-Δ63 his3-Δ200 ade2* GAL$^+$).

Isolation of triterpenes from cell pellets. SCD-LeuHE cultures (10 mL) of SMY8[pRS305Gal], ABY1, and ABY2 were inoculated from single colonies and grown at 30 °C until saturated; these cultures (using 1 mL inocula) were used to start 100-mL SCG-LeuHE cultures. The galactose cultures were removed either in log-phase or once saturated, as determined by checking OD_{600} of each culture. Cultures were harvested by centrifugation; the media was set aside for further analysis. Each cell pellet was saponified in 10 mL 10% KOH in 80% ethanol. BHT (2 mg in ethanol) was added prior

to saponification, as well as the internal standard epicoprostanol (250 μg in ethanol) for quantitation. The mixture was incubated in a 70 °C water bath for 2 h and allowed to cool to room temperature. The NSL were extracted with 3 10-mL aliquots of hexanes, and the combined extracts were washed with 2 5-mL aliquots of distilled water before evaporation of solvent in vacuo. The residues were transferred to vials with small portions of hexanes. Ten percent of each extract was analyzed by GC-MS and GC-FID.

Isolation of triterpene products from media. Hydrophobic resin was used to adsorb any organic materials present in the medium. The media from each culture was transferred to methanol-activated Diaion-20 resin (5 g) and allowed to incubate at 30 °C with shaking overnight. The resin-medium mixture was put through a Kontes column and the medium was allowed to drain. The resin was washed with 3 × 15 mL ethanol and the combined ethanol extracts were then partitioned with 20-mL diH_20 and 3 15-mL aliquots of hexanes. The hexane extracts were washed with 2 × 5 mL diH_2O prior to evaporation of solvent, as described above. Epicoprostanol (250 μg in ethanol) was added to the ethanolic extracts prior to hexane extractions as an internal standard. Ten percent of each extract was analyzed by GC-MS and GC-FID.

TMS-derivatization of extracts. After initial GC-FID and GC-MS analysis, solvents from all samples were evaporated under a nitrogen stream and re-dissolved in 100 μL pyridine and 100 μL BSTFA (bis(trimethylsilyl)trifluoroacetamide). Samples were incubated at 37 °C overnight and then analyzed by GC-FID and GC-MS.

Large-scale production of DOS. Three 1-L cultures were prepared with 10-mL inocula, using YP medium supplemented with 2% adenine and 4% galactose. These cultures were grown to saturation and harvested by centrifugation. The media from each culture was transferred to 15-g methanol-activated resin and incubated overnight at 30 °C, while the cell pellet was saponified and the non-saponifiable lipids extracted with 3 × 100 mL hexanes. The resin was eluted with ethanol and the ethanol partitioned with one-half volume water and 3 × 100 mL hexanes. Epicoprostanol was added as internal standard prior to the extraction. All solvent was removed in vacuo. All samples were analyzed by GC-FID and GC-MS.

DOS purification. DOS was purified from 2-L resin extracts of ABY2 by silica gel chromatography. The extracts were dissolved in 2% ether in hexanes and loaded onto a short column (containing 6 g silica gel). The first fractions, eluted with 2% ether in hexanes, contained squalene. OS was eluted with 4% ether and hexanes. DOS was found in fractions eluted with 6% ether in hexanes. GC-MS analysis of one fraction showed 97% DOS and 3% epicoprostanol.

Quantitation of triterpene standards by GC-FID. Four standards – squalene, OS, DOS, and epicoprostanol – were accurately weighed into glass vials (within 2% error) and left under the vacuum dessicator overnight. The samples were weighed again and dissolved in *p*-xylene to give 1.0 mg/mL stock solutions. Mixtures were prepared from the stock solutions, measuring 100 μL of each standard solution with a gas-tight

Hamilton syringe into a GC vial containing an insert. The six samples were then analyzed by GC-FID and GC-MS.

TMS-derivatization of triterpene standards. After analysis, the GC samples were divided in half, with 100 μL set aside for NMR analysis. The remaining samples (100 μL) were placed under nitrogen streams to remove solvents and were then derivatized overnight (100 μL pyridine + 100 μL BSTFA). The TMS-derivatized samples were then analyzed by GC-FID and GC-MS.

Quantitation of triterpene standards by NMR. Solvent from the remaining non-derivatized samples was evaporated under nitrogen. These samples were then put under vacuum overnight to remove as much solvent as possible and ultimately used for ^{1}H NMR analysis. The 600 MHz NMR was used, with 64 scans per sample. Integration of distinct methyl peaks (Figure 5.6 and Table 5.9) provided ratios.

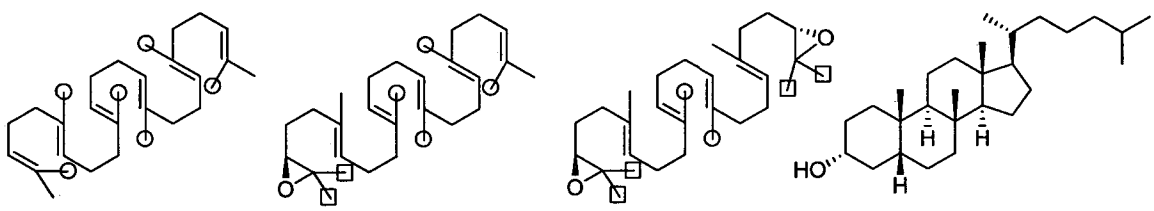

Figure 5.6. Structures of triterpene standards squalene, OS, DOS, and epicoprostanol. Circled methyl groups correspond to 1.60 ppm, and boxed methyl groups correspond to 1.26 and 1.30 ppm.

Table 5.9. ^{1}H NMR chemical shifts for the distinct methyl groups of squalene, OS, and DOS. For 1.26 ppm – 1.68 ppm, the number of methyls corresponding to each chemical shift are noted.

Compound	1.26+1.30 ppm	1.60 ppm	1.62 ppm	1.68 ppm
Squalene	0	6	0	2
OS	1+1	4	1	1
DOS	2+2	2	2	0

To determine the ratios of each compound in the mixtures, the integrated chemical shift areas were compared. However, some chemical shifts corresponded to more than one compound (Table 5.9), so equations were constructed to calculate the contributions of each compound. In all equations, x corresponds to squalene, y to DOS, and e to epicoprostanol. Samples containing epicoprostanol compared the peaks in the table above to the epicoprostanol methyl peaks at 0.63 ppm and 0.92 ppm. Epicoprostanol has no chemical shifts in common with the acyclic compounds so no equations were formulated. From the data collected, only two mixtures could be accurately integrated due to contaminations present in the other samples interfering with the methyl shifts to be integrated. Therefore only the squalene + DOS and DOS + epicoprostanol numbers are reported here.

For the squalene + DOS mixture, the 1.68 ppm shift corresponded only to squalene, while the 1.26 ppm, 1.30 ppm, and 1.62 ppm shifts were characteristic for only DOS (Table 5.10). The two compounds both have 1.60 ppm peaks, so the equation below was utilized to determine the ratios.

$$1.60 \text{ ppm} = 6x + 2y$$

Table 5.10. NMR integration results from the squalene + DOS mixture. Chemical shifts are calculated in ppm and integration values are based on peak areas. Squalene corresponds to x and DOS to y.

Chemical shift	Integration value	Equation	Calculated values
1.26	0.986	2y	y = 0.463
1.30	1.00	2y	y = 0.50
1.60	3.388	6x + 2y	x = 0.410, y = 0.463
1.62	0.886	2y	y = 0.443
1.68	0.893	2x	x = 0.446

From these integration values, squalene was calculated as being 89% of DOS. Comparing 1.26 ppm and 1.30 ppm, there is a 1.4% difference in integration. The allylic and aliphatic methyl groups integrated similarly to one another as well. The correction factor here is thus DOS/squalene = 0.90, with a 1.4% possible error.

Finally, a mixture of DOS and epicoprostanol was examined (Table 5.11). The two methyl groups integrated for epicoprostanol were within 3% of one another. The allylic methyl peaks in DOS integrated similarly to the aliphatic methyl groups, with an 8% difference in integration values. The DOS/epicoprostanol ratio was thus calculated as 0.87 within 11% error.

Table 5.11. NMR integration results from the DOS + epicoprostanol mixture. Chemical shifts are reported in ppm and integration values determined from peak areas. Epicoprostanol corresponds to e and DOS corresponds to y.

Chemical shift	Integration value	Equation	Calculated values
0.64	3	e	e = 3
0.92	2.9	e	e = 2.9
1.26	5.1	2y	y = 2.6
1.30	5.3	2y	y = 2.7
1.60	4.9	2y	y = 2.5
1.62	5.1	2y	y = 2.6

Determination of correction factors for DOS quantitation. The data from both GC-FID quantitation of TMS-derivatized samples and NMR quantitation of the non-derivatized mixtures were taken together to determine correction factors for DOS quantitation (Table 5.12). The NMR values were thought to provide a more accurate picture of the relative amounts of squalene, OS, DOS, or epicoprostanol present in the given mixture; however the results were not definitive. The corrected values are shown in Tables 5.7 and 5.8 in the Results and Discussion section.

Table 5.12. Correction factors calculated from NMR integration and GC-FID quantitation of TMS-derivatized samples. These factors were applied to the SMY8[pRS305Gal], ABY1, and ABY2 quantitation results.

Mixture	Before TMS-derivatization	After TMS-derivatization	NMR calculated ratio	Final correction factor
DOS/Squalene	0.88	0.86	0.90	Negligible
DOS/epicoprostanol	0.854	0.628	0.83	1.32

Chapter 6

Conclusions

Metabolic engineering has come to the forefront in biosynthesizing terpenoids, and the work presented here shows the incredible potential this technology has for increasing terpene production in a microbial host. By manipulating enzymes in the *S. cerevisiae* sterol biosynthetic pathway, terpene precursor accumulation increased, yielding higher levels of terpenoid products. This optimization was achieved by improving activity of foreign terpene synthases or by increasing carbon flux to the desired terpenoid precursors through over-expression of rate-limiting enzymes or deletion of competing enzymes in the *S. cerevisiae* sterol biosynthetic pathway.

We began with an existing engineered strain that overproduces the diterpene precursor geranylgeranyl pyrophosphate (GGPP). Expression of two diterpene synthases from the *Arabidopsis thaliana* gibberellin biosynthetic pathway, *ent*-copalyl pyrophosphate synthase and *ent*-kaurene synthase, led to milligram-level accumulation of the *ent*-copalyl pyrophosphate hydrolysis products *ent*-copalol, *ent*-manool, and *ent*-13-epimanool, as well as the tetracyclic product *ent*-kaurene. Additionally, the GGPP hydrolysis products geranylgeraniol and geranyllinalool were isolated in high yields. This work demonstrated the utility of increasing substrate accumulation to improve terpene production in vivo.

To further enhance yields, a novel method was developed in which a chloroplast processing enzyme from *Arabidopsis* was co-expressed in our diterpene-producing systems. This enzyme cleaved plastid-targeting peptides from the co-expressed diterpene

synthases, forming mature proteins that then exhibited better enzymatic activity. The strain containing *Abies grandis* abietadiene synthase showed a 1.5-fold increase in abietadiene production when the CPE was added, while the strain with *Arabidopsis ent*-copalyl pyrophosphate synthase and *ent*-kaurene synthase saw a 5-to 10-fold increase in *ent*-kaurene production when the CPE was present.

Yeast strains with a non-functional squalene synthase accumulate farnesyl pyrophosphate (FPP), which readily hydrolyzes to farnesol and nerolidol. This work revealed that a majority of the farnesol and nerolidol production is non-enzymatic, and media pH controlled the product profile towards either farnesol or nerolidol. A more acidic pH directed FPP hydrolysis towards nerolidol production, whereas farnesol was formed as the major product under a more neutral pH.

Finally, over-expression of a soluble form of the rate-limiting enzyme HMG-CoA reductase in a strain lacking lanosterol synthase activity accumulated large amounts of the triterpene substrates oxidosqualene and dioxidosqualene, with twenty times more material isolated than that observed in the parent strain. Over-expression of squalene epoxidase facilitates oxidation of oxidosqualene to dioxidosqualene, with 80 times more dioxidosqualene than oxidosqualene accumulating.

This thesis demonstrates several novel approaches to enhancing terpenoid production in microbial hosts. As terpene precursor accumulation was dramatically increased in sesquiterpene, diterpene, and triterpene systems, these strains can be utilized to express terpene synthases and aid in enzyme characterization. Metabolic engineering of microbes has yet to reach optimal levels, and more modifications can still be undertaken to further increase terpenoid production.

Chapter 7

Experimental Procedures

Materials. Media ingredients were purchased from United States Biological (Swampscott, MA). Hemin chloride (heme), ergosterol, Tween 80, bis(trimethylsilyl)trifluoroacetamide (BSTFA), and Triton X-100 were from Sigma-Aldrich (St. Louis, MO). HP Diaion-20 resin was obtained from Supelco (Bellefonte, PA). Solvents, silica gel 60, and TLC plates were obtained from EMD Chemicals, Inc. (Gibbstown, NJ). Restriction enzymes, DNA ladders, and Quick Ligation Kit were purchased from New England Biolabs (Beverly, MA), and the High Purity Plasmid Miniprep System was from Marligen Biosciences, Inc. (Ijamsville, MD). A Qiagen Gel Extraction Kit was used to extract DNA from agarose gel (Valencia, CA). PCR amplification was performed with Triple Master polymerase from Eppendorf (Westbury, NY) and dNTPs from Takara (Fisher Scientific, Pittsburg, PA). The TOPO-TA Cloning Kit for Sequencing and One-Shot chemically competent cells were purchased from Invitrogen (Carlsbad, CA) and used according to the manufacturer's instructions.

Gas chromatography-flame ionization detection (GC-FID). GC-FID analysis was done with an Agilent 6890 GC-FID instrument using a Restek Rtx-35 column (30 m × 0.25 mm × 0.1 μm). Samples (1 μL) were injected into the inlet heated to 280 °C with a 40:1 split ratio. The flame ionization detector temperature was 290 °C, with air flow at 264 mL/min, hydrogen flow at 39 mL/min, and helium makeup gas flow at 37 mL/min. The column was set to constant pressure, dependent on the method used.

For sesquiterpene analysis, a ramping temperature program was used as follows: 50 °C held for 1 min and ramping 15 °C/min until 200 °C was reached. A ramp of 40 °C/min to a final temperature of 280 °C was used, holding at 280 °C for 15 min. The column was held at a constant pressure of 15.13 psi.

Diterpene extracts were analyzed using a similar ramping program. The oven started at 100 °C for 1 min, followed by a 15 °C/min ramp to 200 °C. This temperature was held 5 min before ramping to 280 °C at a rate of 40 °C/min. The final temperature was held 22 min, giving a total run time of 36.67 min. The column and inlet were held at a constant pressure of 6.53 psi.

For triterpene analysis, samples were run on an isothermic program of 260 °C for 32 min, with a constant pressure of 25.73 psi.

Gas chromatography-mass spectrometry (GC-MS). GC-MS analysis was performed on an Agilent 6890N GC instrument interfaced to an Agilent 5973N MSD system with a Restek Rtx-35ms column (30 m × 0.25 mm × 0.1 μm). Aliquots of 1 – 5 μL were injected at a 40:1 split ratio into an inlet at 280 °C. The MSD was heated to 280 °C and set to analyze after a 3 min solvent delay.

For sesquiterpene analysis, chromatography was run with a ramping temperature program, starting at 50 °C for one min and ramping 15 °C/min until it reached 200 °C. This temperature was held for 5 min before increasing 40 °C/min to a final temperature of 280 °C. This final temperature was held for 5 min. The column had a constant pressure of 7.79 psi, with a helium flow of 0.9 mL/min. The MSD was set to analyze 50-500 amu after a 3 min solvent delay.

Analysis of diterpene products utilized a similar ramping program, with the oven starting at 100 °C for 1 min. A 15 °C/min ramp to 200 °C followed, with this temperature held for 5 min. Finally the temperature was increased 40 °C/min to a final temperature of 280 °C, which was held for 15 min. A constant flow of helium travelled through the column at 1 mL/min, with a column pressure of 7.58 psi. The MSD analyzed 50 – 500 amu after a 3 min solvent delay.

For triterpene analysis, chromatography was run on an isothermic program of 260 °C for 32 min. The column was set to a constant flow of 1 mL/min with a pressure of 19.22 psi. The MSD was set to analyze 50-650 amu after a 3 min solvent delay.

Nuclear magnetic resonance (NMR). ^{1}H NMR spectra were acquired at 25 °C in dilute $CDCl_3$ solution on Varian Inova 600 or 800 MHz or Bruker 500 MHz spectrometers. Spectra were referenced to tetramethylsilane at 0 ppm. Deuterated chloroform was filtered over activated basic alumina prior to use and contained a 5:1 mixture of $CDCl_3$ and $CDCl_3$ with 0.05% TMS.

UV-Visible spectroscopy. A Shimadzu UV-visible spectrophotometer UV-1601 was used for all measurements, analyzing at 600 nm wavelength. Disposable cuvettes (10 mM UV-grade) were purchased from VWR International, Inc. (Westchester, PA).

Incubations. All microbial plate cultures were grown in Fisher Scientific Isotemp incubators, set at either 30 °C (for *S. cerevisiae*) or 37 °C (for *E. coli*). Liquid cultures

were grown in either a New Brunswick Series 25 Incubator Shaker or a New Brunswick C24 Incubator Shaker, both set at 200 rpm with temperature set at either 30 °C or 37 °C.

Centrifugation. Centrifugation of Eppendorf tubes (1.5 mL) were performed using either an Eppendorf Centrifuge 5415D or 5810R. Larger tubes (15-mL, 50-mL Falcon tubes or 250-mL centrifuge bottles) were centrifuged using an Eppendorf Centrifuge 5810 at variable speeds and temperatures.

***E. coli* and yeast strains.** For plasmid work, the *E. coli* strain DH5α (F^- *end*A1 *hsd*R17(r_k^-, m_k^+) *sup*E44 *thi*-1 λ^- *rec*A1 *gyr*A96 *rel*A1 Δ(*lac*ZYA-*arg*F)$_{U169}$(m80*lac*ZΔM15))[163] was used. Five yeast strains were utilized for metabolic engineering work. For sesquiterpene production, three different strains were analyzed. PMY1 (MAT**a** *pGAL1-trHMG1::LEU2 erg9::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 his3-Δ200 ade2 GAL*$^+$) is a squalene synthase deletion mutant with a truncated HMG-CoA reductase (trHMG1) over-expressed. Two PMY1 derivatives PMY2 (MAT**a** *dpp1::KanMX pGAL1-trHMG1::LEU2 erg9::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 his3-Δ200 ade2 GAL*$^+$) and BEJY14 (MAT**a** *lpp1::KanMX pGAL1-trHMG1::LEU2 erg9::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 his3-Δ200 ade2 GAL*$^+$)[150] contain DPP1 and LPP1 deletions, respectively. These three strains all accumulate the C-15 linear precursor farnesyl pyrophosphate.

The EHY18 yeast strain (*MAT***a** *pGAL1-BTS1::his pGAL1-trHMG1::LEU2 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2* Gal$^+$)[35,111] was used for diterpene experiments

and contained trHMG1 and GGPP synthase over-expressed, accumulating the C-20 intermediate geranylgeranyl pyrophosphate.

SMY8 (*MAT***a** *erg7::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2* GAL^+)[120] is a lanosterol synthase deletion mutant that accumulates the C-30 triterpene precursors oxidosqualene and dioxidosqualene. This strain was engineered further to increase accumulation of these intermediates.

Bacterial media. *E. coli* cultures were grown in Luria-Bertani broth (LB).[163] This was prepared by dissolving 10 g of LB powder, containing 5 g tryptone, 2.5 g yeast extract, and 2.5 g NaCl, in 500 mL mqH_2O. The LB medium was sterilized by autoclaving at 250 °C for 35 min. Solid LB media was prepared as described above except with the addition of 7.5 g agar/500 mL prior to sterilization. Selective plates were supplemented with the antibiotic ampicillin to a final concentration of 100 μg/mL.

For *E. coli* transformations, cells were incubated with SOC media prior to plating. SOC media[163] contained 2.5 g yeast extract, 10 g tryptone, 1.8 g glucose, 10 mM NaCl (292 mg), 2.5 mM KCl (93 mg), 10 mM $MgCl_2$ (1.017 g anhydrous, 2.03 g hexahydrate), and 20 mM $MgSO_4$ (2.465 g), all dissolved in 500 mL mqH_2O and autoclaved as described above.

Preparation of competent cells also required TB media,[163] prepared by dissolving 1.51 g PIPES, 9.31 g KCl, and 0.832 g $CaCl_2$ in 500 mL mqH_2O. The pH was adjusted to 6.7 with 5 M NaOH. Once everything was dissolved, 5.44 g $MnCl_2 \times 4H_2O$ was added. The mixture was filter-sterilized and stored at 4 °C.

Yeast media. Yeast media were sterilized by autoclaving at 250 °C for 25 or 35 min. Optimal growth required equal volumes of both a nitrogen source and a carbon source. The nitrogen sources used were either YP or synthetic complete. YP was prepared as a 2 × solution with 5 g yeast extract and 10 g peptone dissolved in 500 mL mqH_2O. Synthetic complete media was prepared with 1.7 g yeast nitrogen base, 5 g ammonium sulfate, and 2 g of amino acid mix dissolved in 500 mL mqH_2O, with the final pH adjusted to ~6 with solid NaOH pellets. The amino acid mix contained 10 g leucine and 2 g of the following: alanine, arginine, asparagine, aspartate, cysteine, glutamine, glutamate, glycine, histidine, isoleucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine, adenine, and uracil. Selective media, such as SC-Ura, used an amino acid mix deficient in the specified component.

Sugar solutions of dextrose or galactose were used as carbon sources. Dextrose was prepared as a 2 × solution containing 20 g dextrose dissolved in 500 mL mqH_2O. Galactose, used to induce enzyme expression under a galactose-inducible promoter, was prepared as a 2 × solution containing 20 g galactose dissolved in 500 mL mqH_2O.

For solid yeast media, 7.5 g agar/500 mL was added to the carbon sources prior to sterilization. Plates contained 15 mL each of a carbon source and a nitrogen source.

For yeast strains with nonfunctional sterol biosynthetic pathways, ergosterol was supplemented into the media as a 100 × solution containing 20 μg/mL in EtOH and the detergent Tween 80. These strains also contain a deletion in the heme biosynthetic pathway to mimic anaerobic growth conditions, as yeast will not import sterols under aerobic conditions. This required supplemental hemin as a 100 × solution (65 mg hemin chloride in 25 mL EtOH and 25 mL mqH_2O with 1 mL 1 M NaOH).

Preparation of DH5α competent cells. For all bacterial work, homemade DH5α competent cells[163] were used. The glycerol stock of DH5α was used to inoculate 5 mL LB and incubated overnight at 37 °C with shaking. A 500-mL portion of freshly-prepared SOC media was inoculated with the 5-mL DH5α culture, incubated at room temperature 6 – 8 h and then transferred to a 4 °C refrigerator overnight. The flask was then put back into the room temperature incubator until the OD_{600} had reached approximately 0.6. At this point the cultures were transferred to chilled 250-mL centrifuge bottles and centrifuged 15 min at 3000 rpm, 4 °C.

The supernatant was removed and each set of cells was resuspended in 80 mL ice-cold, freshly-prepared TB media and incubated on ice 10 min. The mixtures were then centrifuged 15 min at 3000 rpm, 4 °C and the supernatant was removed. The cells were resuspended in 10 mL TB per bottle and resuspended with gentle swirling. DMSO was added to a final concentration of 7% (0.75 mL DMSO per 10 mL TB) and the mixture was incubated on ice 10 min. The chilled solutions were transferred to chilled 1.5-mL Eppendorf tubes in 50 μL aliquots. The tubes of cells were frozen by immersion in liquid nitrogen and immediately transferred to a -80 °C freezer for storage.

DNA purification. For mini-prep DNA purification *E. coli* cultures were inoculated in 12-well plates containing 2-mL LB + Amp and incubated overnight at 37 °C. The well plates were centrifuged for 10 min at 3000 rpm, 4 °C and the supernatants were discarded. The cell pellets were resuspended in 100 μL chilled P1 buffer containing Rnase A (50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 20 mg/mL Rnase A). To lyse the

cells, 100 µL P2 buffer (200 mM NaOH, 1% SDS (w/v))was added and the mixture was incubated at room temperature up to 5 min, followed immediately by addition of an equal volume of chilled P3 buffer (3.1 M potassium acetate, pH 5.5). These mixtures incubated at 4 °C for at least 30 min and then centrifuged at 3000 rpm for 10 min. The supernatants were transferred to 1.5-mL Eppendorf tubes and DNA pellets were precipitated with two volumes EtOH and 0.1 M NaCl (600 µL EtOH and 6 µL 5 M NaCl). The tubes left at -20 °C for at least 30 min and centrifuged 20 min at 12,000 rpm, 4 °C. The supernatants were removed by decanting and the DNA pellets were left at 37 °C to allow the remaining EtOH to evaporate. The dried pellets were resuspended in 30 µL Tris pH 8.0 buffer (10 mM Tris-HCl pH 8.0).

For large-scale DNA purification bacterial cultures of either 50 or 100 mL were grown overnight at 37 °C and centrifuged for 20 min at 3500 rpm. The supernatants were removed by decantation and pipette, and the cell pellets were resuspended in 2 mL chilled P1 buffer with Rnase A. This was followed by the addition of 2 mL P2 buffer and incubation on the bench for 5 min. An equal volume of chilled P3 buffer was added and the mixtures were incubated at 4 °C at least 30 min. The tubes were centrifuged for 10 min at 3500 rpm and the supernatants were filtered over Mira Cloth into a clean 50-mL Falcon tube. To this was added 0.7 volumes isopropyl alcohol (IPA), and the mixture was aliquoted into 1.5-mL Eppendorf tubes and incubated at -20 °C for at least 30 min. The tubes were centrifuged for 20 min at 12000 rpm and 4 °C and the supernatants were discarded. The DNA pellets were placed in a 37 °C incubator to remove the residual IPA. To one tube was added 400 µL mqH_2O and the DNA pellet was resuspended by pipetting the mixture up and down several times. This mixture was added to the second tube

containing a DNA pellet, and this procedure was repeated until all DNA pellets were in one tube containing 400 µL water. The tube was incubated at 37 °C to ensure all DNA had dissolved and then centrifuged in a microfuge for 1 min. The supernatant was transferred to a clean Eppi tube and two volumes EtOH and 0.1 M NaCl were added (600 µL EtOH and 6 µL 5 M NaCl). This was mixed by inverting the tube several times and then incubated at -20 °C for at least 30 min. Following centrifugation (20 min at 12,000 rpm and 4 °C), the supernatant was discarded and the DNA pellet was dried at 37 °C. The pellet was ultimately resuspended in 50 µL Tris pH 8.0 buffer.

Alternatively a Marligen High Purity Plasmid Miniprep System was used (Marligen Bioscience, Inc., Ijamsville, MD). The LB + Amp cultures were harvested by centrifugation, as described above. The manufacturer's protocol was followed, except 2 mL each P1 + Rnase A, P2 and P3 were used instead of 400 µL.

DNA sequencing was performed by Lone Star Labs, Inc. (Houston, TX) or by SeqWright DNA Technology Services (Houston, TX).

Polymerase chain reaction (PCR). Oligonucleotides for PCR and sequencing were synthesized by Sigma-Genosys (Houston, TX) or Integrated DNA Technologies (Coralville, IA). Upon arrival, the DNA pellets were centrifuged and dissolved in mqH_2O to a stock concentration of 100 pmol/µL. For PCR, the stock solutions were diluted further to 20 pmol/µL. All were stored at -20 °C.

PCR amplification was carried out on a Mastercycler Gradient system (Eppendorf, Hamburg, Germany). A CAPS program was used, with the following

temperature program: 95 °C for 1 min, 65 °C for 30 s, 72 °C for 3 min, and a 5 min extension at 72 °C, all repeated 36 times.

PCR amplification typically was performed using either plasmid DNA or a cDNA library as template. All reactions contained 0.5 μL template, 20 pmol of each primer, 4 μL dNTPs, 5 μL of the required buffer, and 0.2 – 1 μL polymerase. Reactions were brought up to 50 μL with sterile mqH_2O.

For colony PCR, the reactions contained 20 pmol of each primer, 4 μL dNTPs, 5 μL PC2 buffer, and 0.5 μL Taq polymerase (a generous gift from Prof. Bonnie Bartel) and were brought to 50 μL total volume with sterile mqH_2O. Yeast or *E. coli* colonies were then used to inoculate each reaction.

DNA restriction digestions. All restriction digestions were done according to the New England BioLabs instructions. To map mini-prep DNA, 1 μL plasmid DNA was digested in a 10 μL reaction, containing 1 μL 10 × buffer and 0.2 – 0.3 μL enzyme and 0.1 μL 100 × BSA (if needed) and the remaining volume was mqH_2O. For large-scale DNA preparation 1 μL of plasmid was digested in a mixture containing 2 μL 10 × buffer, 0.5 μL enzyme, 0.2 μL 100 × BSA (if needed), and mqH_2O up to 20 μL total volume.

For cloning and subcloning, between 5 and 20 μL of plasmid DNA was digested in 50 – 100 μL reactions. These included 10 × buffer, 0.5 – 1 μL of each enzyme, 100 × BSA (as needed), and mqH_2O up to the final volume.

For integrative plasmids, linearization was performed prior to yeast transformations. Typically 25 – 50 μL plasmid DNA was digested in 200 μL total reaction volume.

Analytical gel electrophoresis. For analytical purposes, a 1% agarose gel was prepared by dissolving 5 g agarose in 500 mL TAE buffer (50 × TAE stock contained 242 g Tris base, 57.1 mL glacial acetic acid, and 100 mL 0.5 M EDTA dissolved in 1 L mqH_2O) and adding ethidium bromide to the melted agarose (5 μL/100 mL agarose gel). To the DNA samples was added a 10 × loading buffer (20% Ficoll 400, 0.1 M EDTA, 0.25% bromophenol blue, and 0.25% xylene cyanol). Samples were loaded onto the gel and analyzed against the DNA ladder marker λ*Bst*E II (2 μL, New England BioLabs).

Gel purification of DNA. A 1% agarose gel was prepared by mixing 5 g agarose in 500 mL of GTAE buffer (same as TAE buffer with addition of guanosine 2.38 g/10 L buffer) and microwaving until a homogeneous mixture was obtained. Ethidium bromide (5 μL/100 mL) was added and the gel was poured into the mold and allowed to set. The DNA samples were analyzed against the DNA ladder marker λ*Bst*E II (5 μL). The desired bands were excised from the gel with a razor blade and purified using the QIAquick Gel Extraction Kit (Qiagen), following the manufacturer's instructions.

Ligations. DNA ligations were performed using a Quick Ligation Kit (New England BioLabs, Beverly, MA) following the manufacturer's instructions. Each reaction contained 50 ng vector and a 3- to 5-fold molar excess of insert, up to 9 μL total volume. A 10-μL aliquot of 2 × Quick Ligation Buffer and 1 μL Quick Ligase was used for each reaction, for a total volume of 20 μL. Incubations were done at room temperature for 5

min to 1 h, with 1 μL removed prior to addition of ligase and 1 μL removed after incubation to run on an analytical gel.

***E. coli* transformations.** A tube of DH5α competent cells was removed from the -80 °C freezer and allowed to thaw on ice 5 min. Once thawed, 5 μL of the ligation mixture was added to the cells and incubated on ice for 30 min. Following a 30-s heat shock at 42 °C, 500 μL SOC media was added and the mixture was incubated with shaking for 30 min to 1 h. Two aliquots of the transformation, typically 50 μL and 200 μL, were plated on LB plates containing antibiotic and incubated at 37 °C overnight.

Yeast transformations. Yeast transformations were carried out using the lithium acetate method.[119] Cultures were centrifuged at 3000 rpm for 3 min and the supernatant was discarded. The cell pellet was washed twice with 10 mL sterile mqH_2O and centrifuged for 3 min at 3000 rpm after each wash. The pellet was then suspended in 100 μL of sterile deionized water and vortexed. To the yeast suspension was added 10 μL plasmid DNA and 50 μL single-stranded DNA, which had been denatured in boiling water for 5 min. After this mixture was vortexed, 2 mL of yeast transformation buffer (40% polyethylene glycol (PEG), 0.1 M lithium acetate, 10 mM Tris pH 7.5, 1 mM EDTA, and 0.1 M dithiothreitol)[119] was added and the solution was vortexed again to mix. This mixture was incubated at room temperature.

After incubation for 6-18 h, 15 mL sterile mqH_2O was added to the solution, re-suspending the cell pellet, and centrifuged at 3000 rpm for 4 min and the supernatant was decanted. The cell pellet was then washed twice with 10 mL sterile mqH_2O and

centrifuged for 4 min at 3000 rpm to remove any remaining PEG. The cell pellet was resuspended in 500 μL of sterile deionized water and 250 μL was plated on selective media.

For integrative plasmids, in pRS305Gal[120] or pRS305biGal, the plasmid DNA was first linearized with *BstE* II or *BsrG* I and then used for transformation. Digested plasmids and 100 μL denatured ssDNA were used to transform yeast, with the remainder of the procedure following that described above.

Saponification and extraction. Dextrose cultures (10 mL) were inoculated from single yeast colonies and grown to saturation. From these cultures, 100-mL galactose cultures were inoculated (either 1-mL or 10-mL inocula) and grown to log-phase or saturation, as determined by checking OD_{600}. For large-scale cultures, a 1% or 10% inocula was used to start at least 1 L in inducing medium. Cells were harvested by centrifugation and the cell pellets were then saponified in a mixture of 10% KOH in 80% ethanol. To prevent oxidation, 2 mg of the antioxidant butylated hydroxytoluene (BHT, Sigma) was added per 1 g cell pellet before saponification. The mixture was heated in a 70 °C water bath for 2 h and then cooled to room temperature. The non-saponifiable lipids (NSL) were extracted with three aliquots of hexane. The combined extracts were washed with two portions of water and one aliquot of brine before the solvent was evaporated *in vacuo*. The residue was transferred to a small vial with hexane and the remaining solvent was evaporated under a nitrogen gas stream.

Isolation of terpenes from media with hydrophobic resin. Once saturated, yeast cultures were transferred to 250-mL flasks containing methanol-activated hydrophobic resin (Diaion-20, Supelco) and were incubated at 30 °C with shaking overnight. Cultures were then filtered through Kontes columns and the cells and media were discarded. The resin remaining in the column was washed with 3 volumes of ethanol and the combined ethanol was then partitioned between one-half volume of water and three one-half volumes of pentane (for sesquiterpenes) or hexanes (for diterpenes and triterpenes). The pentane or hexane extracts were concentrated by rotary evaporation and transferred to small vials with pentane, hexane, or methylene chloride. Ten percent of each extract was used for GC-MS and GC-FID analysis.

Quantitation of terpene products. To quantitate the terpene products, the internal standard longifolene (for sesquiterpenes) or epicoprostanol (for diterpenes and triterpenes) was added for comparison. To quantitate NSL, the internal standard was added prior to saponification. The internal standard was added to the ethanolic fractions to quantitate resin extracts, prior to extraction with hexane or pentane. Samples were analyzed by GC-MS and GC-FID, and the peak areas corresponding to terpene products were compared to those of the internal standard.

Preparation of TMS-ethers for triterpene analysis. For analysis of triterpene products, it was sometimes necessary to convert the alcohols to TMS-ethers for more accurate quantitation. This was done by transferring a small amount of the sample in a GC vial containing a 300-μL insert and evaporating the solvent under a nitrogen stream.

The samples were then dissolved in a 1:1 mixture of pyridine (dried over KOH) and bis(trimethylsilyl)trifluoroacetamide (BSTFA, Sigma-Aldrich). Samples were mixed by inverting several times, and the vials were left at 37 °C overnight to ensure complete conversion to the TMS-ether.

Column chromatography. To purify terpene products, all column chromatography utilized silica gel 60 (250 – 400 mesh). For sesquiterpenes, gradient systems of pentane/ether were used for separation, while mixtures of hexane/ether were used for diterpenes and triterpenes. Samples were dissolved in a small amount of the same solvent system being used for purification and added to the pre-washed column. Fractions were monitored by TLC.

References

(1) Rohmer, M.; Seemann, M.; Horbach, S.; Bringer-Meyer, S.; Sahm, H. *J. Am. Chem. Soc.* **1996**, *118*, 2564-2566.

(2) Schwender, J.; Seemann, M.; Lichtenthaler, H. K.; Rohmer, M. *Biochem J.* **1996**, *316*, 73-80.

(3) Lange, B. M.; Wildung, M. R.; McCaskill, D.; Croteau, R. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 2100-2104.

(4) Servouse, M.; Mons, N.; Baillargeat, J.-L.; Karst, F. *Biochem. Biophys. Res. Commun.* **1984**, *123*, 424-430.

(5) Hiser, L.; Basso, M. E.; Rine, J. *J. Biol. Chem.* **1994**, *269*, 31383-31389.

(6) Middleton, B.; Tubbs, P. K. *Biochem. J.* **1972**, *126*, 27-34.

(7) Servouse, M.; Karst, F. *Biochem. J.* **1986**, *240*, 541-547.

(8) Schroepfer, G. J. *Annu. Rev. Biochem.* **1981**, *50*, 585-621.

(9) Dietchy, J. M.; Brown, M. S. *J. Lipid Res.* **1974**, *15*, 508-514.

(10) Basson, M. E.; Thorsness, M.; Rine, J. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 5563-5567.

(11) Liscum, L.; Finer-Moore, J.; Stroud, R. M.; Luskey, K. L.; Brown, M. S.; Goldstein, J. L. *J. Biol. Chem.* **1985**, *260*, 522-530.

(12) Goldstein, J. L.; Brown, M. S. *Nature* **1990**, *343*, 425-430.

(13) Dimster-Denk, D.; Thorsness, M. K.; Rine, J. *Mol. Biol. Cell* **1994**, *5*, 655-665.

(14) Hampton, R. Y.; Rine, J. *J. Cell Biol.* **1994**, *125*, 299-312.

(15) Hampton, R. Y.; Bhakta, H. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12944-12948.

(16) Gardner, R. G.; Hampton, R. Y. *J. Biol. Chem.* **1999**, *274*, 31671-31678.

(17) Wright, R.; Basson, M.; D'Ari, L.; Rine, J. *J. Cell. Biol.* **1988**, *107*, 101-114.

(18) Polakowski, T.; Stahl, U.; Lang, C. *Appl. Microbiol. Biotechnol.* **1998**, *49*, 66-71.

(19) Donald, K. A. G.; Hampton, R. Y.; Fritz, I. B. *Appl. Environ. Microbiol.* **1997**, *63*, 3341-3344.

(20) Karst, F.; Lacroute, F. *Mol. Gen. Genetics* **1977**, *154*, 269-277.

(21) Tsay, Y. H.; Robinson, G. W. *Mol. Cell. Biol.* **1991**, *11*, 620-631.

(22) Toth, M. J.; Huwyler, L. *J. Biol. Chem.* **1996**, *271*, 7895-7898.

(23) Chambon, C.; Ladeveze, V.; Servouse, M.; Blanchard, L.; Javelot, C.; Vladescu, B.; Karst, F. *Lipids* **1991**, *26*, 633-636.

(24) Anderson, M. S.; Muehlbacher, M.; Street, I. P.; Proffitt, J.; Poulter, C. D. *J. Biol. Chem.* **1989**, *264*, 19169-19175.

(25) Ashby, M. N.; Edwards, P. A. *J. Biol. Chem.* **1990**, *265*, 13157-13164.

(26) Davis, E. M.; Croteau, R. *Top. Curr. Chem.* **2000**, *209*, 53-95.

(27) Anderson, M. S.; Yarger, J. G.; Burck, C. L.; Poulter, C. D. *J. Biol. Chem.* **1989**, *264*, 19176-19184.

(28) Chambon, C.; Ladeveze, V.; Oulmouden, A.; Servouse, M.; Karst, F. *Curr. Genet.* **1990**, *18*, 41-46.

(29) Blanchard, L.; Karst, F. *Gene* **1993**, *125*, 185-189.

(30) Fernandez, S. M. S.; Kellogg, B. A.; Poulter, C. D. *Biochemistry* **2000**, *39*, 15316-15321.

(31) Karst, F.; Plochocka, D.; Meyer, S.; Szkopinska, A. *Cell Biol. Int.* **2004**, *28*, 193-197.

(32) Szkopinska, A.; Swiezewska, E.; Karst, F. *Biochem. Biophys. Res. Commun.* **2000**, *267*, 473-477.

(33) Lorenz, R. T.; Rodriguez, R. J.; Lewis, T. A.; Parks, L. W. *J. Bacteriol.* **1986**, *167*, 981-985.

(34) Jiang, Y.; Proteau, P.; Poulter, D.; Ferro-Novick, S. *J. Biol. Chem.* **1995**, *270*, 21793-21799.

(35) Hart, E. A.; Ph. D. Thesis, Rice University, 2001.

(36) Clarke, S. *Annu. Rev. Biochem.* **1992**, *61*, 355-386.

(37) Moores, S. L.; Schaber, M. D.; Mosser, S. D.; Rands, E.; O'Hara, M. B.; Garsky, V. M.; Marshall, M. S.; Pompliano, D. L.; Gibbs, J. B. *J. Biol. Chem.* **1991**, *266*, 14603-14610.

(38) Sasiak, K.; Rilling, H. C. *Arch. Biochem. Biophys.* **1988**, *260*, 622-627.

(39) Jennings, S. M.; Tsay, Y. H.; Fisch, T. M.; Robinson, G. W. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 6038-6042.

(40) Rilling, H. C.; Poulter, C. D.; Epstein, W. W.; Larsen, B. *J. Am. Chem. Soc.* **1971**, *93*, 1783-1785.

(41) Grabowska, D.; Karst, F.; Szkopinska, A. *FEBS Lett.* **1998**, *434*, 406-408.

(42) Song, L. *Anal. Biochem.* **1993**, *317*, 180-185.

(43) Tchen, T. T.; Bloch, K. *J. Am. Chem. Soc.* **1955**, *77*, 6085-6086.

(44) Jandrositz, A.; Turnowsky, F.; Högenauer, G. *Gene* **1991**, *107*, 155-160.

(45) M'Baya, B.; Karst, F. *Biochem. Biophys. Res. Commun.* **1987**, *147*, 556-564.

(46) Tchen, T. T.; Bloch, K. *J. Biol. Chem.* **1957**, *226*, 931-939.

(47) Klein, H. P. *J. Bacteriol.* **1955**, *69*, 620-627.

(48) Leber, R.; Zenz, R.; Schrottner, K.; Fuchsbichler, S.; Puhringer, B.; Turnowsky, F. *Eur. J. Biochem.* **2001**, *268*, 914-924.

(49) M'Baya, B.; Fegueur, M.; Servouse, M.; Karst, F. *Lipids* **1989**, *24*, 1020-1023.

(50) Bai, M.; Xiao, X. Y.; Prestwich, G. D. *Biochem. Biophys. Res. Commun.* **1992**, *185*, 323-329.

(51) Veen, M.; Stahl, U.; Lang, C. *FEMS Yeast Res.* **2003**, *4*, 87-95.

(52) Corey, E. J.; Matsuda, S. P. T.; Bartel, B. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 2211-2215.

(53) Corey, E. J.; Gross, S. K. *J. Am. Chem. Soc.* **1967**, *89*, 4561-4562.

(54) Boutaud, O.; Dolis, D.; Schuber, F. *Biochem. Biophys. Res. Commun.* **1992**, *188*, 898-904.

(55) Gardner, R. G.; Shan, H.; Matsuda, S. P. T.; Hampton, R. Y. *J. Biol. Chem.* **2001**, *276*, 8681-8694.

(56) Kalb, V. F.; Woods, C. W.; Turi, T. G.; Dey, C. R.; Sutter, T. R.; Loper, J. C. *DNA Cell Biol.* **1987**, *6*, 529-537.

(57) Marcireau, C.; Guyonnet, D.; Karst, F. *Curr. Genet.* **1992**, *22*, 267-272.

(58) Lorenz, R. T.; Parks, L. W. *DNA Cell Biol.* **1992**, *11*, 685-692.

(59) Bard, M.; Bruner, D. A.; Pierson, C. A.; Lees, N. D.; Biermann, B.; Frye, L.; Koegel, C.; Barbuch, R. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 186-190.

(60) Gachotte, D.; Barbuch, R.; Gaylor, J.; Nickel, E.; Bard, M. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 13794-13799.

(61) Gachotte, D.; Sen, S. E.; Eckstein, J.; Barbuch, R.; Krieger, M.; Ray, B. D.; Bard, M. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 12655-12660.

(62) Gachotte, D.; Eckstein, J.; Barbuch, R.; Hughes, T.; Roberts, C.; Bard, M. *J. Lipid Res.* **2001**, *42*, 150-154.

(63) Gaber, R. F.; Copple, D. M.; Kennedy, B. K.; Vidal, M.; Bard, M. *Mol. Cell. Biol.* **1989**, *9*, 3447-3456.

(64) Arthington, B. A.; Hoskins, J.; Skatrud, P. L.; Bard, M. *Gene* **1991**, *107*, 173-174.

(65) Ashman, W. H.; Barbuch, R. J.; Ulbright, C. E.; Jarrett, H. W.; Bard, M. *Lipids* **1991**, *26*, 628-632.

(66) Arthington, B. A.; Bennett, L. G.; Skatrud, P. L.; Guynn, C. J.; Barbuch, R. J.; Ulbright, C. E.; Bard, M. *Gene* **1991**, *102*, 39-44.

(67) Skaggs, B. A.; Alexander, J. F.; Pierson, C. A.; Schweitzer, K. S.; Chun, K. T.; Koegel, C.; Barbuch, R.; Bard, M. *Gene* **1996**, *169*, 105-109.

(68) Zweytick, D.; Hrastnik, C.; Kohlwein, S. D.; Daum, G. *FEBS Lett.* **2000**, *470*, 83-87.

(69) Todd, R. B.; Andrianopolous, A. *Fungal Genet. Biol.* **1997**, *21*, 388-405.

(70) Crowley, J. H.; Leak, F. W.; Jr.; Shianna, K. V.; Tove, S.; Parks, L. W. *J. Bacteriol.* **1998**, *180*, 4177-4183.

(71) Lewis, T. L.; Keesler, G. A.; Fenner, G. P.; Parks, L. W. *Yeast* **1988**, *4*, 93-106.

(72) Leak, F. W.; Tove, S.; Parks, L. W. *DNA Cell Biol.* **1999**, *18*, 133-139.

(73) Corey, E. J.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **1991**, *113*, 8172-8174.

(74) Toke, D. A.; Bennett, W. L.; Dillon, D. A.; Wu, W.; Chen, X.; Ostrander, D. B.; Oshiro, J.; Cremesti, A.; Voelker, D. R.; Fischl, A. S.; Carman, G. M. *J. Biol. Chem.* **1998**, *273*, 3278-3284.

(75) Toke, D. A.; Bennett, W. L.; Oshiro, J.; Wu, W.; Voelker, D. R.; Carman, G. M. *J. Biol. Chem.* **1998**, *273*, 14331-14338.

(76) Faulkner, A.; Chen, X.; Rush, J.; Horazdovsky, B.; Waechter, C. J.; Carman, G. M.; Sternweis, P. C. *J. Biol. Chem.* **1999**, *274*, 14831-14837.

(77) Song, L. *Appl. Biochem. Biotech.* **2006**, *128*, 149-157.

(78) Fray, R. G.; Wallace, A.; Fraser, P. D.; Valero, D.; Hedden, P.; Bramley, P. M.; Grierson, D. *Plant J.* **1995**, *8*, 693-701.

(79) Besumbes, O.; Sauret-Gueto, S.; Phillips, M. A.; Imperial, S.; Rodriguez-Concepcion, M.; Boronat, A. *Biotechnol. Bioeng.* **2004**, *88*, 168-175.

(80) Oswald, M.; Fischer, M.; Dirninger, N.; Karst, F. *FEMS Yeast Res.* **2007**, *7*, 413-421.

(81) Reiling, K. K.; Yoshikuni, Y.; Martin, V. J. J.; Newman, J. D.; Bohlmann, J.; Keasling, J. D. *Biotechnol. Bioeng.* **2004**, *87*, 199-212.

(82) Jackson, B. E.; Hart-Wells, E. A.; Matsuda, S. P. T. *Org. Lett.* **2003**, *5*, 1629-1632.

(83) Chang, M. C. Y.; Eachus, R. A.; Trieu, W.; Ro, D.-K.; Keasling, J. D. *Nature Chem. Biol.* **2007**, *3*, 274-277.

(84) Takahashi, S.; Yeo, Y.; Greenhagen, B. T.; McMullin, T.; Song, L.; Maurina-Brunker, J.; Rosson, R.; Noel, J. P.; Chappell, J. *Biotechnol. Bioeng.* **2007**, *97*, 170-181.

(85) Asadollahi, M. A.; Maury, J.; Moller, K.; Nielsen, K. F.; Schalk, M.; Clark, A.; Nielsen, J. *Biotechnol. Bioeng.* **2008**, *99*, 666-677.

(86) Huang, Q. L.; Roessner, C. A.; Croteau, R.; Scott, A. I. *Bioorg. Med. Chem. Lett.* **2001**, *9*, 2237-2242.

(87) Cyr, A.; Wilderman, P. R.; Determan, M.; Peters, R. J. *J. Am. Chem. Soc.* **2007**, *129*, 6684-6685.

(88) Ro, D.-K.; Paradise, E. M.; Ouellet, M.; Fisher, K. J.; Newman, K. L.; Ndungu, J. M.; Ho, K. A.; Eachus, R. A.; Ham, T. S.; Kirby, J.; Chang, M. C. Y.; Withers, S. T.; Shiba, Y.; Sarpong, R.; Keasling, J. D. *Nature* **2006**, *440*, 940-943.

(89) DeJong, J. M.; Liu, Y.; Bollon, A. P.; Long, R. M.; Jennewein, S.; Williams, D.; Croteau, R. B. *Biotechnol. Bioeng.* **2006**, *93*, 212-224.

(90) Bohlmann, J.; Meyer-Gauen, G.; Croteau, R. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 4126-4133.

(91) Richter, S.; Lamppa, G. K. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 7463-7468.

(92) LaFever, R. E.; Stofer Vogel, B.; Croteau, R. *Arch. Biochem. Biophys.* **1994**, *313*, 139-149.

(93) Williams, D. C.; Wildung, M. R.; Jin, A. Q.; Dalal, D.; Oliver, J. S.; Coates, R. M.; Croteau, R. *Arch. Biochem. Biophys.* **2000**, *379*, 137-146.

(94) Huang, K. X.; Huang, Q. L.; Wildung, M. R.; Croteau, R.; Scott, A. I. *Protein Expr. Purif.* **1998**, *13*, 90-96.

(95) Peters, R. J.; Flory, J. E.; Jetter, R.; Ravn, M. M.; Lee, H.-J.; Coates, R. M.; Croteau, R. B. *Biochemistry* **2000**, *39*, 15592 -15602.

(96) Prisic, S.; Peters, R. J. *Plant Physiol.* **2007**, *144*, 445-454.

(97) Corey, E. J.; Matsuda, S. P. T.; Bartel, B. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11628-11632.

(98) Kolesnikova, M. D.; Xiong, Q.; Lodeiro, S.; Hua, L.; Matsuda, S. P. T. *Arch. Biochem. Biophys.* **2006**, *447*, 87-95.

(99) Kolesnikova, M. D.; Obermeyer, A. C.; Wilson, W. K.; Lynch, D. A.; Xiong, Q.; Matsuda, S. P. T. *Org. Lett.* **2007**, *9*, 2183-2186.

(100) Kolesnikova, M. D.; Wilson, W. K.; Lynch, D. A.; Obermeyer, A. C.; Matsuda, S. P. T. *Org. Lett.* **2007**, *9*, 5223-5226.

(101) Xiong, Q.; Wilson, W. K.; Matsuda, S. P. T. *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 1285-1288.

(102) Lodeiro, S.; Xiong, Q.; Wilson, W. K.; Kolesnikova, M. D.; Onak, C.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2007**, *129*, 11213-11222.

(103) Shan, H.; Segura, M. J. R.; Wilson, W. K.; Lodeiro, S.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2005**, *127*, 18008-18009.

(104) Kirby, J.; Romanini, D. W.; Paradise, E. M.; Keasling, J. D. *FEBS J.* **2008**, *275*, 1852-1859.

(105) Hedden, P.; Phillips, A. L. *Trends Plant Sci.* **2000**, *5*, 523-530.

(106) Wani, M. C.; Taylor, H. L.; Wall, M. E.; Coggon, P.; McPhail, A. T. *J. Am. Chem. Soc.*. **1971**, *93*, 2325-2327.

(107) Robinson, D. R.; West, C. A. *Biochemistry* **1970**, *9*, 70-79.

(108) Robinson, D. R.; West, C. A. *Biochemistry* **1970**, *9*, 80-89.

(109) Hill, A. M.; Cane, D. E.; Mau, J. D.; West, C. A. *Arch. Biochem. Biophys.* **1996**, *336*, 283-289.

(110) Jennewein, S.; Park, H.; DeJong, J. M.; Long, R. M.; Bollon, A. P.; Croteau, R. B. *Biotechnol. Bioeng.* **2005**, *89*, 588-598.

(111) Matsuda, S. P. T.; Hart, E. A. *U. S. Patent 7,238,514*, 2004.

(112) Hedden, P.; Kamiya, Y. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1997**, *48*, 431-460.

(113) Sun, T. P.; Kamiya, Y. *Plant Cell* **1994**, *6*, 1509-1518.

(114) Yamaguchi, S.; Sun, T. P. K., H.; Kamiya, Y. *Plant Physiol.* **1998**, *116*, 1271-1278.

(115) Yee, N. K. N.; Coates, R. M. *J. Org. Chem.* **1992**, *57*, 4598-4608.

(116) Barrero, A. F.; Sanchez, J. F.; Alvarez-Manzaneda, E. J.; Munoz Dorado, M.; Haidour, A. *Phytochemistry* **1993**, *32*, 1261-1265.

(117) Paradise, E. M.; Kirby, J.; Chan, R.; Keasling, J. D. *Biotechnol. Bioeng.* **2008**, *100*, 371-378.

(118) Kleinig, H. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1989**, *40*, 39-59.

(119) Schiestl, R. H.; Gietz, R. D. *Curr. Genet.* **1989**, *16*, 339-346.

(120) Corey, E. J.; Matsuda, S. P. T.; Baker, C. H.; Ting, A. Y.; Cheng, H. *Biochem. Biophys. Res. Commun.* **1996**, *219*, 327-331.

(121) Gonzalez, A. G.; Fraga, B. N.; Hernandez, M. G.; Hanson, J. R. *Phytochemistry* **1981**, *20*, 846-847.

(122) Dudley, M. W.; Dueber, M. T.; West, C. A. *Plant Physiol.* **1986**, *81*, 335-342.

(123) Moore, T. C.; Coolbaugh, R. C. *Phytochemistry* **1976**, *15*, 1241-1247.

(124) Railton, I. D.; Fellows, B.; West, C. A. *Phytochemistry* **1984**, *23*, 1261-1267.

(125) Smith, M. W.; Yamaguchi, S.; Ait-Ali, T.; Kamiya, Y. *Plant Physiol.* **1998**, *118*, 1411-1419.

(126) Stofer Vogel, B.; Wildung, M. R.; Vogel, G.; Croteau, R. *J. Biol. Chem.* **1996**, *271*, 23262-23268.

(127) Mau, C. J.; West, C. A. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 8497-8501.

(128) von Heijne, G.; Stepphun, J.; Herrmann, R. G. *Eur. J. Biochem.* **1989**, *180*, 535-545.

(129) Keegstra, K.; Olsen, L. J. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1989**, *40*, 471-501.

(130) Richter, S.; Lamppa, G. K. *J. Cell. Biol.* **1999**, *147*, 33-43.

(131) Takahashi, S.; Yeo, Y.-S.; Zhao, Y.; O'Maille, P. E.; Greenhagen, B. T.; Noel, J. P.; Coates, R. M.; Chappell, J. *J. Biol. Chem.* **2007**, *282*, 31744-31754.

(132) Millis, J. R.; Saucy, G. G.; Maurina-Brunker, J.; McMullin, T. W.; PCT. Int. Appl., 2000.

(133) Whitesides, G. M.; Lewis, D. W. *J. Am. Chem. Soc.* **1970**, *92*, 6979-6980.

(134) Wilson, W. K.; Scallen, T. J.; Morrow, C. J. *J. Lipid Res.* **1982**, *23*, 645-652.

(135) Bouwmeester, H. J.; Verstappen, F. W. A.; Posthumus, M. A.; Dicke, M. *Plant Physiol.* **1999**, *121*, 173-180.

(136) Degenhardt, J.; Gershenzon, J. *Planta* **2000**, *210*, 815-822.

(137) Schnee, C.; Kollner, T. G.; Gershenzon, J.; Degenhardt, J. *Plant Physiol.* **2002**, *130*, 2049-2060.

(138) Poulter, C. D.; Rilling, H. C. *Biochemistry* **1976**, *15*, 1079-1083.

(139) Carrau, F. M.; Medina, K.; Boido, E.; Farina, L.; Gaggero, C.; Dellacassa, E.; Versini, G.; Henschke, P. A. *FEMS Microbiol. Lett.* **2005**, *243*, 107-115.

(140) Benedict, C. R.; Lu, J.-L.; Pettigrew, D. W.; Liu, J.; Stipanovic, R. D.; Williams, H. J. *Plant Physiol.* **2001**, *125*, 1755-1765.

(141) Farnesal mass spectrum obtained from Wiley Subscription Services, Inc. through SciFinder Scholar, 2008.

(142) Cornwell, P. A.; Barry, B. W. *J. Pharm. Pharmacol.* **1994**, *46*, 261-269.

(143) Yamane, M. A.; Williams, A. C.; Barry, B. W. *J. Pharm. Pharmacol.* **1995**, *47*, 978-989.

(144) Kubo, I.; Muroi, H.; Himejima, M. *J. Agric. Food Chem.* **1992**, *40*, 245-248.

(145) Kubo, I.; Morimitsu, Y. *J. Agric. Food Chem.* **1995**, *43*, 1626-1628.

(146) Lopes, N. P.; Kato, M. J.; Andrade, E. H.; Maia, J. G.; Yoshida, M.; Plancart, A. R.; Katzin, A. M. *J. Ethnopharmacol.* **1999**, *67*, 313-319.

(147) Arruda, D. C.; D'Alexandri, F. L.; Katzin, A. M.; Uliana, S. R. B. *Antimicrob. Agents Chemother.* **2005**, *49*, 1679-1687.

(148) Lee, S. J.; Han, J. I.; Lee, G. S.; Park, M. J.; Choi, I. G.; Na, K. J.; Jeung, E. B. *Biol. Pharm. Bull.* **2007**, *30*, 184-188.

(149) Lindahl, A.-L.; Olsson, M. E.; Mercke, P.; Tollbom, O.; Schelin, J.; Brodelius, M.; Brodelius, P. E. *Biotechnology Lett.* **2005**, *28*, 571-580.

(150) Jackson, B. E.; Ph. D. Thesis, Rice University, 2004.

(151) Schwartz, M. A.; Swanson, G. C. *J. Org. Chem.* **1979**, *44*, 953-958.

(152) Xu, R.; Fazio, G. C.; Matsuda, S. P. T. *Phytochemistry* **2004**, *65*, 261-290.

(153) Fazio, G. C.; Xu, R.; Matsuda, S. P. T. *J. Am. Chem. Soc.*. **2004**, *126*, 5678-5679.

(154) Adams, B. G. *J. Bacteriol.* **1972**, *111*, 308-315.

(155) Fazio, G. C.; Ph. D. Thesis, Rice University, 2006.

(156) De Pascual Teresa, J.; Urones, J. G.; Marcos, I. S.; Basabe, P.; Sexmero Cuadrado, M. J.; Fernandez Moro, R. *Phytochemistry* **1987**, *26*, 1767-1776.

(157) Della Greca, M.; Fiorentino, A.; Monaco, P.; Previtera, L. *Phytochemistry* **1994**, *35*, 1017-1022.

(158) Xiang, T.; Shibuya, M.; Katsube, Y.; Tsutsumi, T.; Otsuka, M.; Zhang, H.; Masuda, K.; Ebizuka, Y. *Org. Lett.* **2006**, *8*, 2835-2838.

(159) Xiang, T.; Shibuya, M.; Katsube, Y.; Tsutsumi, T.; Otsuka, M.; Zhang, H.; Masuda, K.; Ebizuka, Y. *Org. Lett.* **2006**, *8*, 2835-2838.

(160) Panini, S. R.; Sexton, R. C.; Gupta, A. K.; Parish, E. J.; Chitrakorn, S.; Rudney, H. *J. Lipid Res.* **1986**, *27*, 1190-1204.

(161) Casey, W. M.; Burgess, J. P.; Parks, L. W. *Biochim. Biophys. Acta* **1991**, *1081*, 279-284.

(162) Parish, E. J.; Parish, S. C.; Li, S. *Crit. Rev. Biochem. Mol. Biol.* **1999**, *34*, 265-272.

(163) Ausubel, F. M.; Brent, R.; Kingston, R. E.; Moore, D. D.; Seidman, J. G.; Smith, J. A.; Struhl, K., Eds. *Current Protocols in Molecular Biology*; Wiley-Interscience: New York, 1999.

Appendix A – List of abbreviations

A. grandis	*Abies grandis*
AgAS	*Abies grandis* abietadiene synthase
Ala	alanine
Amp	ampicillin
Asn	asparagine
*At*GA1	*Arabidopsis thaliana ent*-copalyl pyrophosphate synthase
*At*GA2	*Arabidopsis thaliana ent*-kaurene synthase
ATP	adenosine 5'-triphosphate
BHT	butylated hydroxytoluene
BiGal	bi-directional galactose promoter
BSA	bovine serum albumin
BSTFA	bis(trimethylsilyl)trifluoroacetamide
$CDCl_3$	deuterated chloroform
cDNA	complementary deoxyribonucleic acid
COSYDEC	decoupled ^1H-^1H correlation NMR spectroscopy
CPE	chloroplast processing enzyme
D	dextrose
DEPT	distortionless enhancement by polarization transfer
diH_2O	distilled water
DMAPP	dimethylallyl pyrophosphate
DMSO	dimethyl sulfoxide

DNA	deoxyribonucleic acid
dNTPs	deoxynucleosides-5'-triphosphates
DOS	dioxidosqualene
DPP1	diacylglycerol pyrophosphate phosphatase
E. coli	*Escherichia coli*
E1	unimolecular elimination
E2	bimolecular elimination
EDTA	ethylene diamine tetraacetic acid
Epi	epicoprostanol
ER	endoplasmic reticulum
EtOH	ethanol
$Eu(hfc)_3$	europium tris[2-(heptafluoropropylhydroxymethylene)-(+)-camphorate]
FAD	flavin adenine dinucleotide
FPP	farnesyl pyrophosphate
G	galactose
GC-FID	gas chromatography-flame ionization detection
GC-MS	gas chromatography-mass spectrometry
GGPP	geranylgeranyl pyrophosphate
Gln	glutamine
Glu	glutamic acid, glutamate
Gly	glycine
GPP	geranyl pyrophosphate
GTAE	40 mM Tris-base, 20 mM acetic acid, 1 mM EDTA, 1 mM guanosine

Heme	hemin chloride
HMBC	heteronuclear multiple bond correlation
HMG-CoA	3-hydroxy-3-methylglutaryl-coenzyme A
IPP	isopentenyl pyrophosphate
KanMX	geneticin-resistant marker
LB	Luria-Bertani
Leu	leucine
LPP1	lipid phosphate phosphatase
m/z	mass-to-charge ratio
MOPS	3-[N-morpholino]propanesulfonic acid
mqH_2O	milli-Q filtered water
mRNA	messenger ribonucleic acid
NAD(P)H	nicotinamide adenine dinucleotide (phosphate)
NCBI	National Center for Biotechnology Information
NMR	nuclear magnetic resonance
NSL	non-saponifiable lipids
OD_{600}	optical density measured at 600 nm
OS	oxidosqualene
OSC	oxidosqualene cyclase
PCR	polymerase chain reaction
PEG	polyethylene glycol
Phe	phenylalanine
PIPES	1,4-piperazinethanesulfonic acid

PtDS	*Pisum sativa* putative diterpene synthase
RnaseA	ribonuclease A
RT-PCR	reverse transcriptase polymerase chain reaction
S. cerevisiae	*Saccharomyces cerevisiae*
SC	synthetic complete
SDS	sodium dodecyl sulfate
Ser	serine
S_N1	unimolecular substitution
S_N2	bimolecular substitution
SrGA1	*Stevia rebaudiana ent*-copalyl pyrophosphate synthase
SrGA2	*Stevia rebaudiana ent*-kaurene synthase
ssDNA	single-stranded DNA
TAE	40 mM Tris-base, 20 mM acetic acid, 1 mM EDTA
TbTS	*Taxus brevifolia* taxadiene synthase
TLC	thin-layer chromatography
TMS	tetramethylsilane (for NMR) or trimethylsilyl (for derivatization)
TMS-epi	TMS-derivatized epicoprostanol
trHMG1	truncated hydroxymethylglutaryl-coenzyme A reductase
Tris-HCl	tris(hydroxymethyl)aminomethane hydrochloride
Trp	tryptophan
U or Ura	uracil
UV	ultraviolet
YP	yeast extract plus peptone

Appendix B – Relevant *S. cerevisiae* enzymes

BTS1	geranylgeranyl pyrophosphate synthase
DPP1	diacylglycerol pyrophosphate phosphatase
ERG1	squalene epoxidase
ERG10	acetoacetyl-CoA thiolase
ERG12	mevalonate kinase
ERG13	HMG-CoA synthase
ERG19	mevalonate-5-pyrophosphate decarboxylase
ERG20	farnesyl pyrophosphate synthase
ERG7	lanosterol synthase
ERG8	phosphomevalonate kinase
ERG9	squalene synthase
HMGR	HMG-CoA reducase
IDI1	isopentenyl pyrophosphate isomerase
LPP1	lipid phosphate phosphatase

Appendix C – List of Relevant Plasmids and Yeast Strains

Plasmids

pCVP2.1	*A. thaliana ent*-copalyl pyrophosphate synthase (GA1) in pRS426GalR subcloned at *Sal* I – *Not* I
pCVP6.1	*A. thaliana ent*-kaurene synthase (GA2) in pRS313biGal subcloned at *Xba* I – *BamH* I
pCVP26.1	*A. thaliana* GA1 and GA2 in pRS426biGal subcloned at *Sac* II – *Not* I (AtGA1) and *Xba* I – *BamH* I (AtGA2)
pAMB5.3	*A. thaliana* chloroplast processing enzyme in pRS313Gal subcloned at *Sal* I – *Not* I
pEH9.0	*Abies grandis* abietadiene synthase in pRS426Gal subcloned at *BamH* I – *Not* I
pEH12.4	*S. cerevisiae* truncated HMG-CoA reductase (trHMG1) in pRS305biGal subcloned at *BamH* I – *Not* I
pGCF6.0	*S. cerevisiae* squalene epoxidase (ERG1) in pRS305Gal subcloned at *Sal* I – *Not* I

Yeast strains

EHY18	*upc2-1* incorporated, trHMG1 under control of GAL1 promoter, GGPP synthase under control of GAL1 promoter (*MAT***a** *pGAL1-BTS1::his pGAL1-trHMG1::LEU2 upc2-1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2* Gal^+)

CPY1	*upc2-1* incorporated, trHMG1 under control of GAL1 promoter, GGPP synthase under control of GAL1 promoter, *At*GA1 under control of Gal10 promoter (*MAT***a** *pGAL10-AtGA1::URA3 pGAL1-BTS1::his pGAL1-trHMG1::LEU2 upc2-1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2* Gal^+)
CPY2	*upc2-1* incorporated, trHMG1 under control of GAL1 promoter, GGPP synthase under control of GAL1 promoter, *At*GA1 under control of Gal10 promoter, *At*GA2 under control of GAL1 promoter (*MAT***a** *pGAL10-AtGA1::URA3 pGAL1-AtGA2::HIS3 pGAL1-BTS1::his pGAL1-trHMG1::LEU2 upc2-1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2* Gal^+)
PMY1	heme auxotroph, squalene synthase deletion, trHMG1 under control of GAL1 promoter (*MAT***a** *pGAL1-trHMG1::LEU2 erg9::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2* Gal^+)
PMY2	heme auxotroph, squalene synthase deletion, trHMG1 under control of GAL1 promoter, (*MAT***a** *pGAL1-trHMG1::LEU2 erg9::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2* Gal^+)
BEJY14	heme auxotroph, squalene synthase deletion, trHMG1 under control of GAL1 promoter,

(*MAT***a** *pGAL1-trHMG1::LEU2 erg9::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2* Gal^+)

SMY8 heme auxotroph, lanosterol synthase deletion (*MAT***a** *erg7::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2* Gal^+)

ABY1 heme auxotroph, lanosterol synthase deletion, trHMG1 under control of GAL1 promoter (*MAT***a** *pGAL1-trHMG1::LEU2 erg7::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2* Gal^+)

ABY2 heme auxotroph, lanosterol synthase deletion, *Sc*ERG1 under control of GAL1 promoter (*MAT***a** *pGAL1-ERG1::LEU2 erg7::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2* Gal^+)

Appendix D – Spectral data

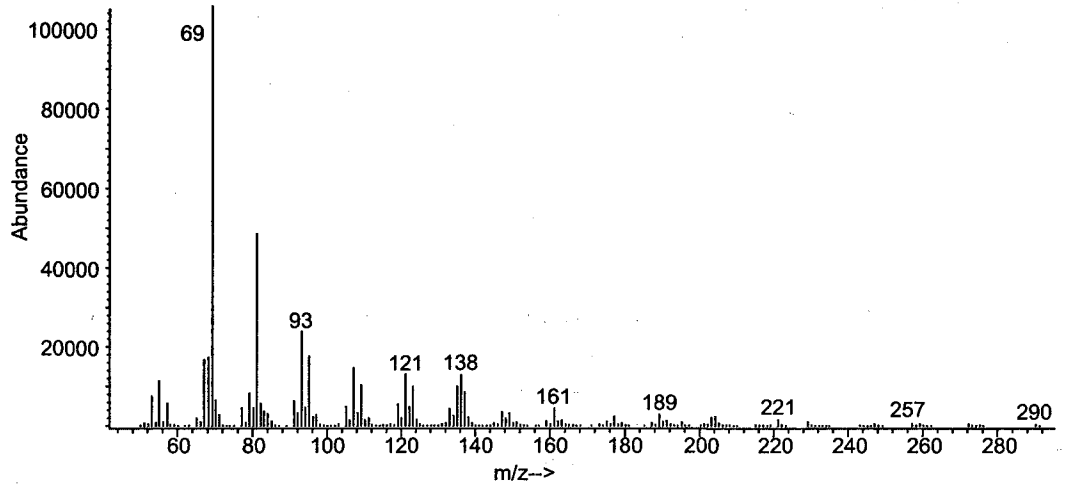

Figure D.1. Electron impact mass spectrum (EI-MS) of geranylgeraniol isolated from EHY18[pRS426Gal] crude NSL. The mass spectral pattern matches that of an authentic geranylgeraniol standard.

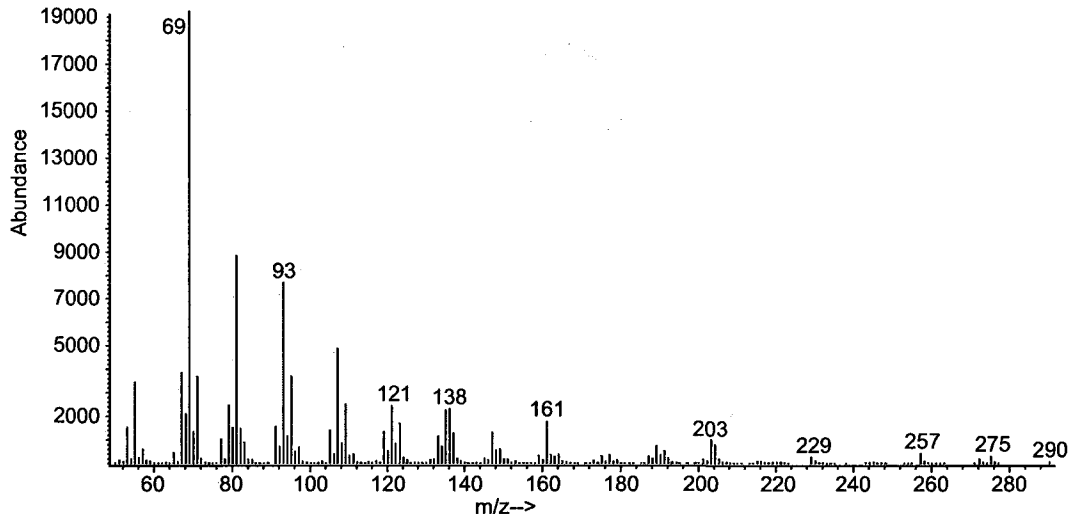

Figure D.2. EI-MS of geranyllinalool isolated from an EHY18[pRS426Gal] crude NSL. The mass spectral pattern matches that of an authentic geranyllinalool standard.

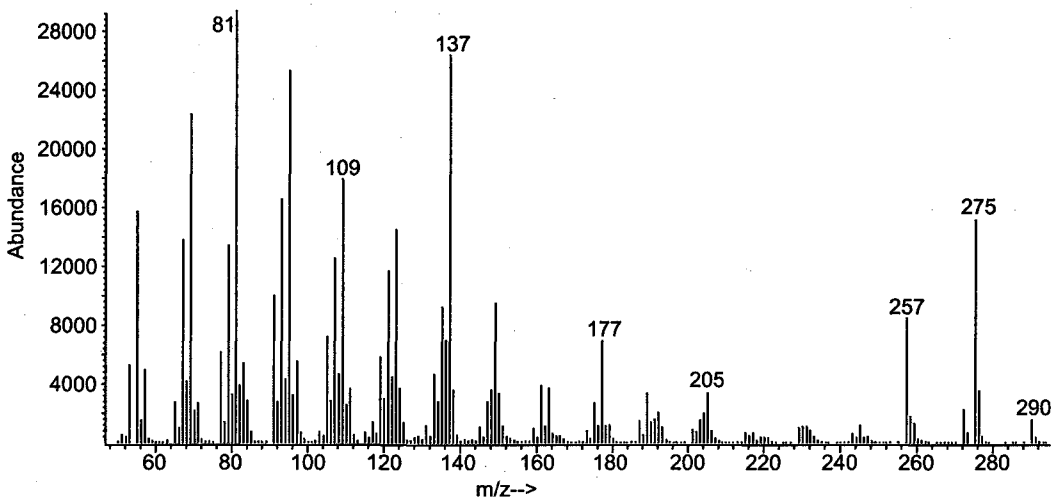

Figure D.3. Mass spectrum of *ent*-copalol from a crude NSL of CPY1.

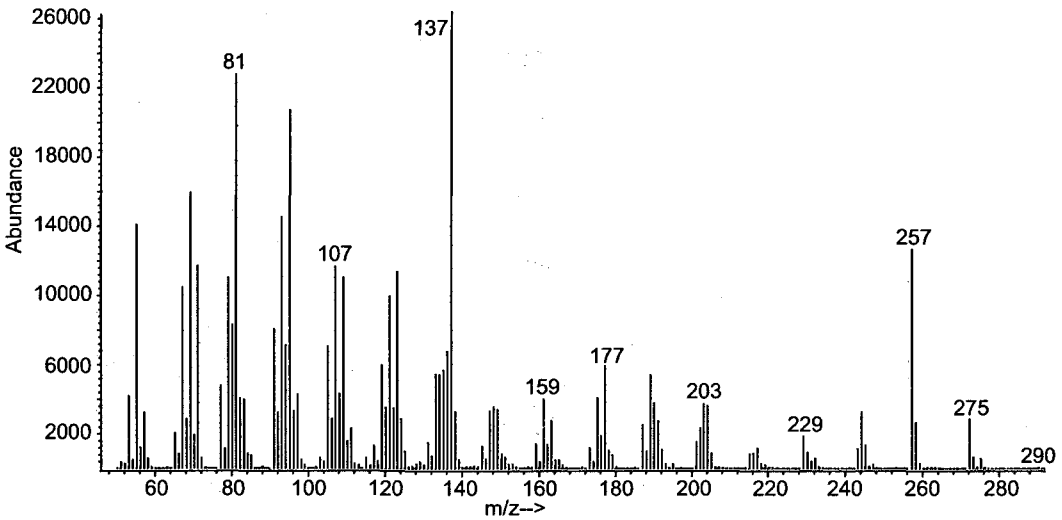

Figure D.4. Mass spectrum of *ent*-manool/*ent*-13-epimanool from a CPY1 crude NSL.

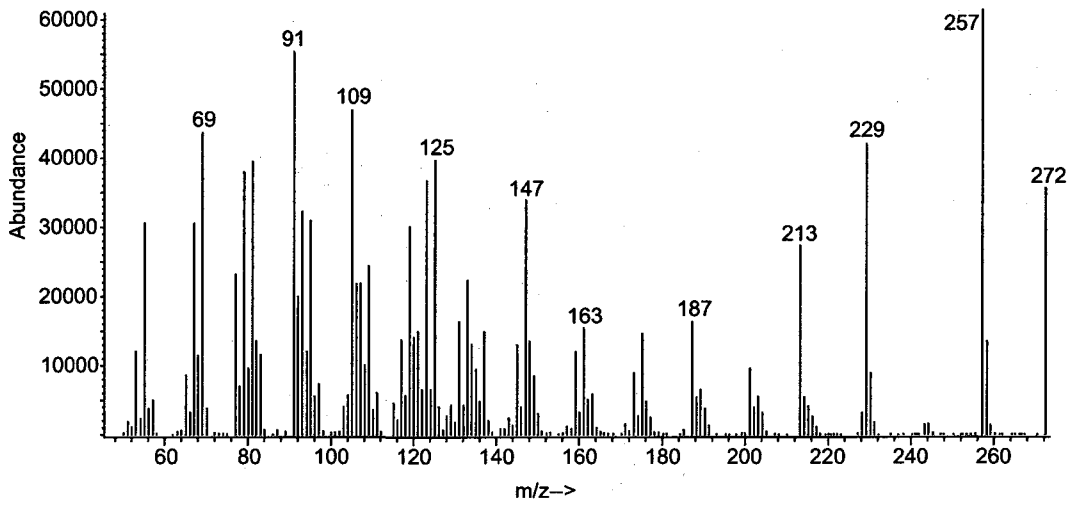

Figure D.5. EI-MS of *ent*-kaurene isolated from a CPY2 crude NSL.

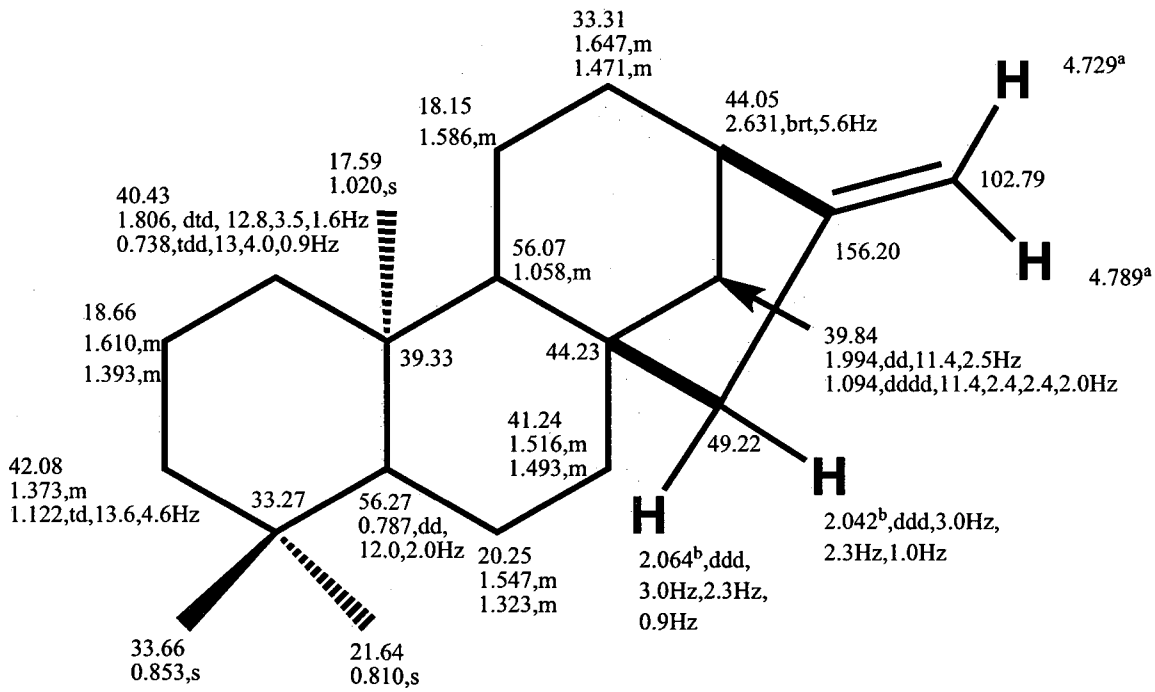

Figure D.6. ^{1}H and ^{13}C NMR chemical shift assignments for *ent*-kaurene.

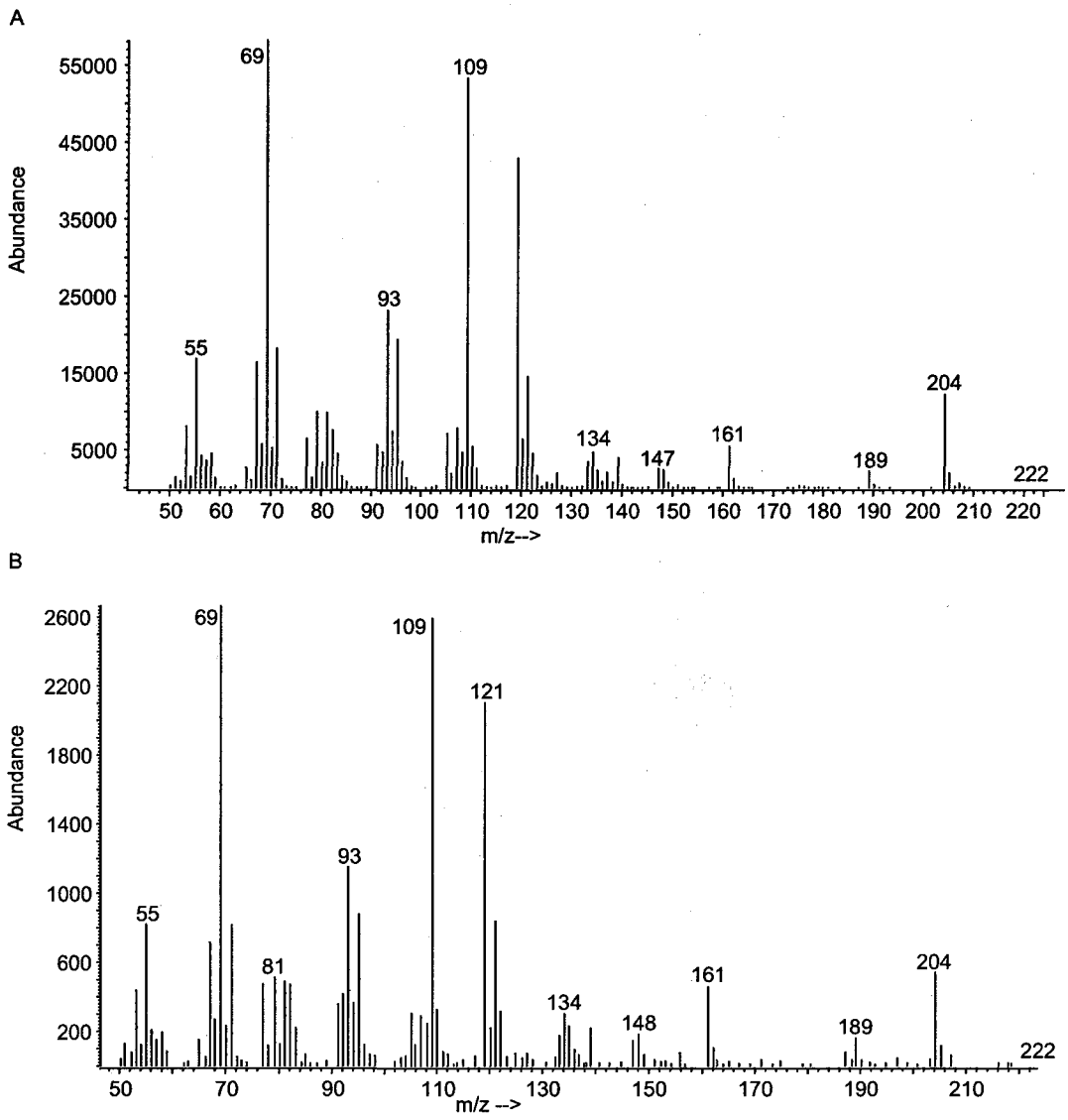

Figure D.7. EI-MS of α-bisabolol. Spectrum A is an authentic standard, and spectrum B is α-bisabolol from PMY1[pRS316Gal].

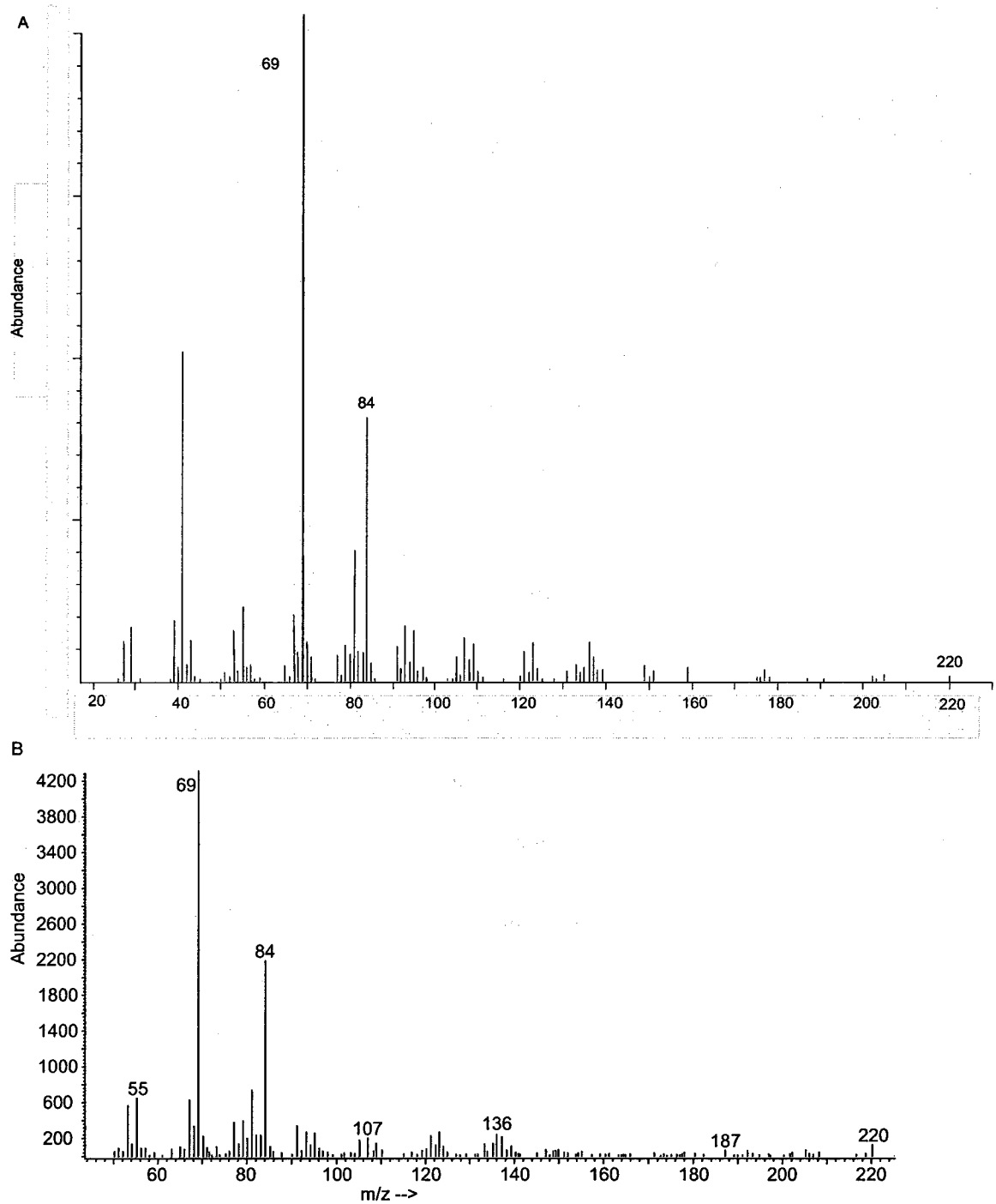

Figure D.8. Farnesal mass spectrum. Spectrum A is from literature, and spectrum B was isolated from PMY1[pRS316Gal].

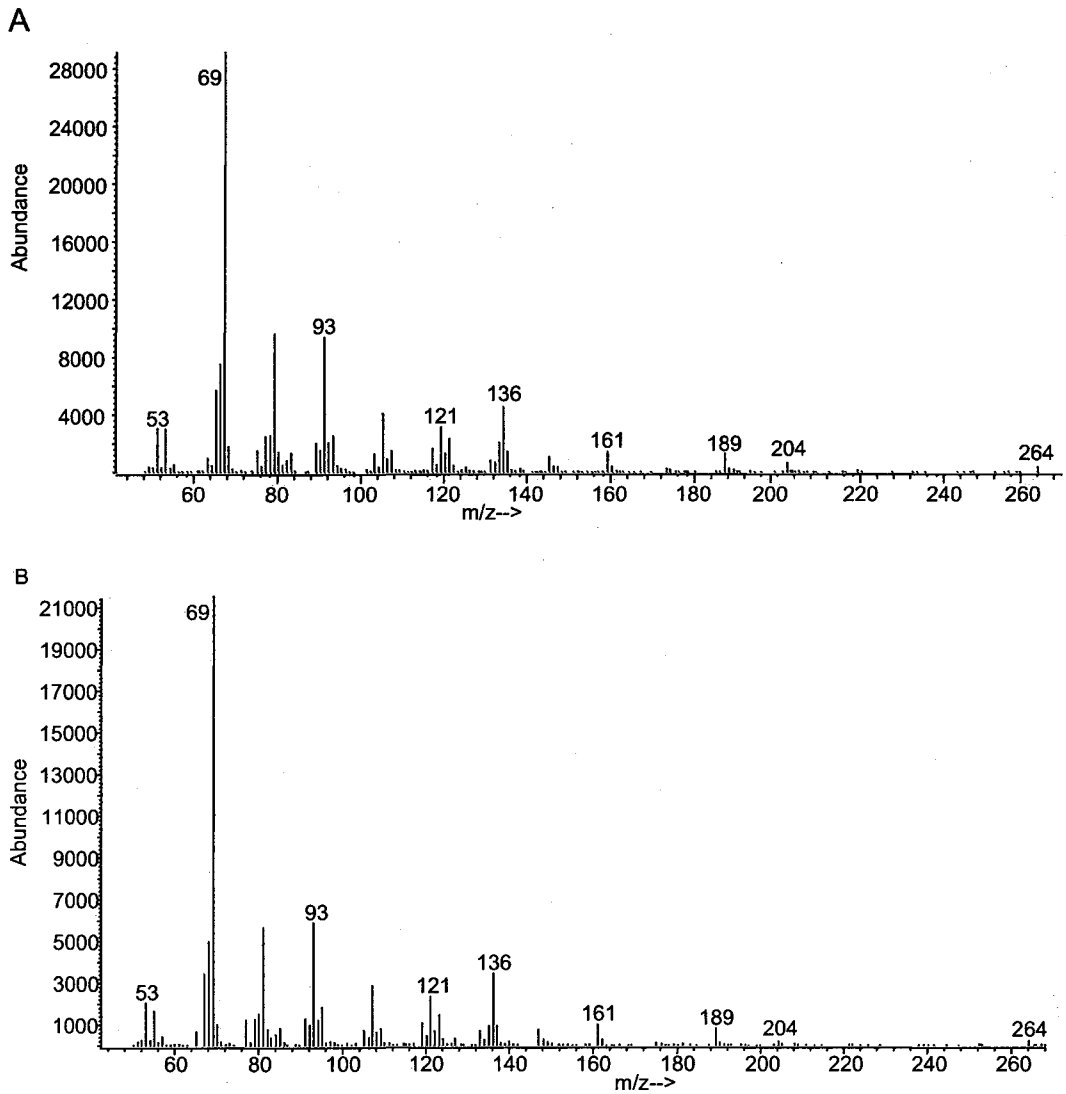

Figure D.9. Mass spectrum of farnesyl acetate. Spectrum A is that of an authentic standard and spectrum B is from PMY1[pRS316Gal].